

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification<sup>6</sup>:</b> <b>C12N 15/55, 9/22, 15/90, 15/85</b>		<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 98/10077</b> <b>(43) International Publication Date:</b> 12 March 1998 (12.03.98)
<b>(21) International Application Number:</b> PCT/US97/15941		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
<b>(22) International Filing Date:</b> 9 September 1997 (09.09.97)			
<b>(30) Priority Data:</b> 08/814,877 9 September 1996 (09.09.96) US 08/850,880 2 May 1997 (02.05.97) US			
<b>(71) Applicant:</b> WISCONSIN ALUMNI RESEARCH FOUNDATION [US/US]; 614 North Walnut Street, P.O. Box 7365, Madison, WI 53707-7365 (US).		<b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
<b>(72) Inventors:</b> REZNIKOFF, William, S.; 733 Lakewood Boulevard, Mapel Bluff, WI 53704 (US). GORYSHIN, Igor, Yu; Apartment D, 23 University Houses, Madison, WI 53705 (US). ZHOU, Hong; Apartment 104, 5331 Brody Drive, Madison, WI 53705 (US).			
<b>(74) Agent:</b> BERSON, Bennett, J.; Quarles & Brady, P.O. Box 2113, Madison, WI 53701-2113 (US).			

**(54) Title:** SYSTEM FOR *IN VITRO* TRANSPOSITION USING MODIFIED TNS TRANSPONASE**(57) Abstract**

A system for *in vitro* transposition includes a donor DNA that includes a transposable element flanked by a pair of bacterial transposon Tn5 outside end repeat sequences, a target DNA into which the transposable element can transpose, and a modified Tn5 transposase having higher binding avidity to the outside end repeat sequences and being less likely to assume an inactive multimer form than wild type Tn5 transposase.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

5

**SYSTEM FOR *IN VITRO* TRANSPOSITION USING MODIFIED TNS TRANSPOSAE****CROSS-REFERENCE TO RELATED APPLICATION**

This patent application is a continuation-in-part of a patent application entitled "System for *In Vitro* Transposition," filed March 11, 1997, for which no serial number has yet been accorded. Applicants have petitioned for a filing date of September 9, 1996 to be accorded to the parent application.

**STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT**

Not applicable.

15

**BACKGROUND OF THE INVENTION**

The present invention relates generally to the field of transposable nucleic acid and, more particularly to production and use of a modified transposase enzyme in a system for introducing genetic changes to nucleic acid.

Transposable genetic elements are DNA sequences, found in a wide variety of prokaryotic and eukaryotic organisms, that can move or transpose from one position to another position in a genome. *In vivo*, intra-chromosomal transpositions as well as transpositions between chromosomal and non-chromosomal genetic material are known. In several systems, transposition is known to be under the control of a transposase enzyme that is typically encoded by the transposable element. The genetic structures and transposition mechanisms of various transposable elements are summarized, for example, in "Transposable Genetic Elements" in "The Encyclopedia of Molecular Biology," Kendrew and Lawrence, Eds., Blackwell Science, Ltd., Oxford (1994), incorporated herein by reference.

*In vitro* transposition systems that utilize the particular transposable elements of bacteriophage Mu and bacterial transposon Tn10 have been described, by the research groups of

5 Kiyoshi Mizuuchi and Nancy Kleckner, respectively.

The bacteriophage Mu system was first described by Mizuuchi, K., "In Vitro Transposition of Bacteria Phage Mu: A Biochemical Approach to a Novel Replication Reaction," Cell:785-794 (1983) and Craigie, R. et al., "A Defined System for the DNA Strand-Transfer Reaction at the Initiation of Bacteriophage Mu Transposition: Protein and DNA Substrate Requirements," P.N.A.S. U.S.A. 82:7570-7574 (1985). The DNA donor substrate (mini-Mu) for Mu *in vitro* reaction normally requires six Mu transposase binding sites (three of about 30 bp at each end) and an enhancer sequence located about 1 kb from the left end. The donor plasmid must be supercoiled. Proteins required are Mu-encoded A and B proteins and host-encoded HU and IHF proteins. Lavoie, B.D, and G. Chaconas, "Transposition of phage Mu DNA," Curr. Topics Microbiol. Immunol. 204:83-99 (1995). The Mu-based system is disfavored for *in vitro* transposition system applications because the Mu termini are complex and sophisticated and because transposition requires additional proteins above and beyond the transposase.

The Tn10 system was described by Morisato, D. and N. Kleckner, "Tn10 Transposition and Circle Formation *in vitro*," Cell 51:101-111 (1987) and by Benjamin, H. W. and N. Kleckner, "Excision Of Tn10 from the Donor Site During Transposition Occurs By Flush Double-Strand Cleavages at the Transposon Termini," P.N.A.S. U.S.A. 89:4648-4652 (1992). The Tn10 system involves the a supercoiled circular DNA molecule carrying the transposable element (or a linear DNA molecule plus *E. coli* IHF protein). The transposable element is defined by complex 42 bp terminal sequences with IHF binding site adjacent to the inverted repeat. In fact, even longer (81 bp) ends of Tn10 were used in reported experiments. Sakai, J. et al., "Identification and Characterization of Pre-Cleavage Synaptic Complex that is an Early Intermediate in Tn10 transposition," E.M.B.O. J. 14:4374-4383 (1995). In the Tn10 system, chemical treatment of the transposase protein is essential to support active transposition. In addition, the termini of the Tn10 element limit its utility in a generalized *in vitro*

5 transposition system.

Both the Mu- and Tn10-based *in vitro* transposition systems are further limited in that they are active only on covalently closed circular, supercoiled DNA targets. What is desired is a more broadly applicable *in vitro* transposition system that utilizes shorter, more well defined termini and which is active on target DNA of any structure (linear, relaxed circular, and supercoiled circular DNA).

#### BRIEF SUMMARY OF THE INVENTION

The present invention is summarized in that an *in vitro* transposition system comprises a preparation of a suitably modified transposase of bacterial transposon Tn5, a donor DNA molecule that includes a transposable element, a target DNA molecule into which the transposable element can transpose, all provided in a suitable reaction buffer.

The transposable element of the donor DNA molecule is characterized as a transposable DNA sequence of interest, the DNA sequence of interest being flanked at its 5'- and 3'-ends by short repeat sequences that are acted upon *in trans* by Tn5 transposase.

The invention is further summarized in that the suitably modified transposase enzyme comprises two classes of differences from wild type Tn5 transposase, where each class has a separate measurable effect upon the overall transposition activity of the enzyme and where a greater effect is observed when both modifications are present. The suitably modified enzyme both (1) binds to the repeat sequences of the donor DNA with greater avidity than wild type Tn5 transposase ("class (1) mutation") and (2) is less likely than the wild type protein to assume an inactive multimeric form ("class (2) mutation"). A suitably modified Tn5 transposase of the present invention that contains both class (1) and class (2) modifications induces at least about 100-fold ( $\pm 10\%$ ) more transposition than the wild type enzyme, when tested in combination in an *in vivo* conjugation assay as described by Weinreich, M.D., "Evidence that the *cis* Preference of the Tn5 Transposase is Caused by

5 Nonproductive Multimerization," Genes and Development 8:2363-  
2374 (1994), incorporated herein by reference. Under optimal  
conditions, transposition using the modified transposase may be  
higher. A modified transposase containing only a class (1)  
mutation binds to the repeat sequences with sufficiently  
10 greater avidity than the wild type Tn5 transposase that such a  
Tn5 transposase induces about 5- to 50-fold more transposition  
than the wild type enzyme, when measured *in vivo*. A modified  
transposase containing only a class (2) mutation is  
15 sufficiently less likely than the wild type Tn5 transposase to  
assume the multimeric form that such a Tn5 transposase also  
induces about 5- to 50-fold more transposition than the wild  
type enzyme, when measured *in vivo*.

In another aspect, the invention is summarized in that a  
method for transposing the transposable element from the donor  
20 DNA into the target DNA *in vitro* includes the steps of mixing  
together the suitably modified Tn5 transposase protein, the  
donor DNA, and the target DNA in a suitable reaction buffer,  
allowing the enzyme to bind to the flanking repeat sequences of  
the donor DNA at a temperature greater than 0°C, but no higher  
25 than about 28°C, and then raising the temperature to  
physiological temperature (about 37°C) whereupon cleavage and  
strand transfer can occur.

It is an object of the present invention to provide a  
useful *in vitro* transposition system having few structural  
30 requirements and high efficiency.

It is another object of the present invention to provide a  
method that can be broadly applied in various ways, such as to  
create absolute defective mutants, to provide selective markers  
35 to target DNA, to provide portable regions of homology to a  
target DNA, to facilitate insertion of specialized DNA  
sequences into target DNA, to provide primer binding sites or  
tags for DNA sequencing, to facilitate production of genetic  
fusions for gene expression studies and protein domain mapping,  
as well as to bring together other desired combinations of DNA  
40 sequences (combinatorial genetics).

It is a feature of the present invention that the modified

5 transposase enzyme binds more tightly to DNA than does wild type Tn5 transposase.

It is an advantage of the present invention that the modified transposase facilitates *in vitro* transposition reaction rates of at least about 100-fold higher than can be achieved using wild type transposase (as measured *in vivo*). It is noted that the wild-type Tn5 transposase shows no detectable *in vitro* activity in the system of the present invention. Thus, while it is difficult to calculate an upper limit to the increase in activity, it is clear that hundreds, if not thousands, of colonies are observed when the products of *in vitro* transposition are assayed *in vivo*.

It is another advantage of the present invention that *in vitro* transposition using this system can utilize donor DNA and target DNA that is circular or linear.

20 It is yet another advantage of the present invention that *in vitro* transposition using this system requires no outside high energy source and no other protein other than the modified transposase.

25 Other objects, features, and advantages of the present invention will become apparent upon consideration of the following detailed description.

#### BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

Fig. 1 depicts test plasmid pRZTL1, used herein to demonstrate transposition *in vitro* of a transposable element located between a pair of Tn5 outside end termini. Plasmid pRZTL1 is also shown and described in SEQ ID NO:3.

35 Fig. 2 depicts an electrophoretic analysis of plasmid pRZTL1 before and after *in vitro* transposition. Data obtained using both circular and linear plasmid substrates are shown.

Fig. 3 is an electrophoretic analysis of plasmid pRZTL1 after *in vitro* transposition, including further analysis of the molecular species obtained using circular and linear plasmid substrates.

40 Fig. 4 shows plasmids pRZ1496, pRZ5451 and pRZTL1, which are detailed in the specification.

5       Fig. 5 shows a plot of papillae per colony over time for  
various mutant OE sequences tested *in vivo* against EK54/MA56  
transposase.

10      Fig. 6 shows a plot of papillae per colony over time for  
various mutant OE sequences with a smaller Y-axis than is shown  
in Fig. 5 tested against EK54/MA56 transposase.

15      Fig. 7 shows a plot of papillae per colony over time for  
various mutant OE sequences tested against MA56 Tn5  
transposase.

20      Fig. 8 shows *in vivo* transposition using two preferred  
mutants, tested against MA56 and EK54/MA56 transposase.

#### DETAILED DESCRIPTION OF THE INVENTION

25      It will be appreciated that this technique provides a  
simple, *in vitro* system for introducing any transposable  
element from a donor DNA into a target DNA. It is generally  
accepted and understood that Tn5 transposition requires only a  
pair of OE termini, located to either side of the transposable  
element. These OE termini are generally thought to be 18 or 19  
bases in length and are inverted repeats relative to one  
another. Johnson, R. C., and W. S. Reznikoff, Nature 304:280  
35      (1983), incorporated herein by reference. The Tn5 inverted  
repeat sequences, which are referred to as "termini" even  
though they need not be at the termini of the donor DNA  
molecule, are well known and understood.

40      Apart from the need to flank the desired transposable  
element with standard Tn5 outside end ("OE") termini, few other  
requirements on either the donor DNA or the target DNA are  
envisioned. It is thought that Tn5 has few, if any,  
preferences for insertion sites, so it is possible to use the  
system to introduce desired sequences at random into target  
DNA. Therefore, it is believed that this method, employing the  
modified transposase described herein and a simple donor DNA,  
is broadly applicable to introduce changes into any target DNA,  
without regard to its nucleotide sequence. It will, thus, be  
applied to many problems of interest to those skilled in the  
art of molecular biology.

5        In the method, the modified transposase protein is  
combined in a suitable reaction buffer with the donor DNA and  
the target DNA. A suitable reaction buffer permits the  
transposition reaction to occur. A preferred, but not  
necessarily optimized, buffer contains spermidine to condense  
10      the DNA, glutamate, and magnesium, as well as a detergent,  
which is preferably 3-[(3-cholamidopropyl) dimethyl-ammonio]-1-  
propane sulfonate ("CHAPS"). The mixture can be incubated at a  
temperature greater than 0°C and as high as about 28°C to  
facilitate binding of the enzyme to the OE termini. Under the  
15      buffer conditions used by the inventors in the Examples, a  
pretreatment temperature of 30°C was not adequate. A preferred  
temperature range is between 16°C and 28°C. A most preferred  
pretreatment temperature is about 20°C. Under different buffer  
conditions, however, it may be possible to use other below-  
20      physiological temperatures for the binding step. After a short  
pretreatment period of time (which has not been optimized, but  
which may be as little as 30 minutes or as much as 2 hours, and  
is typically 1 hour), the reaction mixture is diluted with 2  
volumes of a suitable reaction buffer and shifted to  
25      physiological conditions for several more hours (say 2-3 hours)  
to permit cleavage and strand transfer to occur. A temperature  
of 37°C, or thereabouts, is adequate. After about 3 hours, the  
rate of transposition decreases markedly. The reaction can be  
stopped by phenol-chloroform extraction and can then be  
30      desalted by ethanol precipitation.

When the DNA has been purified using conventional  
purification tools, it is possible to employ simpler reaction  
conditions in the *in vitro* transposition method. DNA of  
sufficiently high purity can be prepared by passing the DNA  
35      preparation through a resin of the type now commonly used in  
the molecular biology laboratory, such as the Qiagen resin of  
the Qiagen plasmid purification kit (Catalog No. 12162). When  
such higher quality DNA is employed, CHAPS can be omitted from  
the reaction buffer. When CHAPS is eliminated from the  
40      reaction buffer, the reactants need not be diluted in the  
manner described above. Also, the low temperature incubation

5 step noted above can be eliminated in favor of a single incubation for cleavage and strand transfer at physiological conditions. A three hour incubation at 37°C is sufficient.

10 Following the reaction and subsequent extraction steps, transposition can be assayed by introducing the nucleic acid reaction products into suitable bacterial host cells (e.g., *E. coli* K-12 DH5 $\alpha$  cells (*recA* $^{-}$ ); commercially available from Life Technologies (Gibco-BRL)) preferably by electroporation, described by Dower et al., *Nuc. Acids. Res.* 16:6127 (1988), and monitoring for evidence of transposition, as is described 15 elsewhere herein.

Those persons skilled in the art will appreciate that apart from the changes noted herein, the transposition reaction can proceed under much the same conditions as would be found in an *in vivo* reaction. Yet, the modified transposase described 20 herein so increases the level of transposition activity that it is now possible to carry out this reaction *in vitro* where this has not previously been possible. The rates of reaction are even greater when the modified transposase is coupled with an optimized buffer and temperature conditions noted herein.

25 In another aspect, the present invention is a preparation of a modified Tn5 transposase enzyme that differs from wild type Tn5 transposase in that it (1) binds to the repeat sequences of the donor DNA with greater avidity than wild type Tn5 transposase and (2) is less likely than the wild type protein to assume an inactive multimeric form. An enzyme having these requirements can be obtained from a bacterial host 30 cell containing an expressible gene for the modified enzyme that is under the control of a promoter active in the host cell. Genetic material that encodes the modified Tn5 transposase can be introduced (e.g., by electroporation) into suitable bacterial host cells capable of supporting expression 35 of the genetic material. Known methods for overproducing and preparing other Tn5 transposase mutants are suitably employed. For example, Weinreich, M. D., et al., *supra*, describes a 40 suitable method for overproducing a Tn5 transposase. A second method for purifying Tn5 transposase was described in de la

5 Cruz, N. B., et al., "Characterization of the Tn5 Transposase  
and Inhibitor Proteins: A Model for the Inhibition of  
Transposition," J. Bact. 175:6932-6938 (1993), also  
incorporated herein by reference. It is noted that induction  
can be carried out at temperatures below 37°C, which is the  
10 temperature used by de la Cruz, et al. Temperatures at least in  
the range of 33 to 37°C are suitable. The inventors have  
determined that the method for preparing the modified  
transposase of the present invention is not critical to success  
15 of the method, as various preparation strategies have been used  
with equal success.

Alternatively, the protein can be chemically synthesized,  
in a manner known to the art, using the amino acid sequence  
attached hereto as SEQ ID NO:2 as a guide. It is also possible  
20 to prepare a genetic construct that encodes the modified  
protein (and associated transcription and translation signals)  
by using standard recombinant DNA methods familiar to molecular  
biologists. The genetic material useful for preparing such  
constructs can be obtained from existing Tn5 constructs, or can  
25 be prepared using known methods for introducing mutations into  
genetic material (e.g., random mutagenesis PCR or site-directed  
mutagenesis) or some combination of both methods. The genetic  
sequence that encodes the protein shown in SEQ ID NO:2 is set  
forth in SEQ ID NO:1.

30 The nucleic acid and amino acid sequence of wild type Tn5  
transposase are known and published. N.C.B.I. Accession Number  
U00004 L19385, incorporated herein by reference.

In a preferred embodiment, the improved avidity of the  
modified transposase for the repeat sequences for OE termini  
35 (class (1) mutation) can be achieved by providing a lysine  
residue at amino acid 54, which is glutamic acid in wild type  
Tn5 transposase. The mutation strongly alters the preference  
of the transposase for OE termini, as opposed to inside end  
("IE") termini. The higher binding of this mutation, known as  
40 EK54, to OE termini results in a transposition rate that is  
about 10-fold higher than is seen with wild type transposase.  
A similar change at position 54 to valine (mutant EV54) also

5 results in somewhat increased binding/transposition for OE termini, as does a threonine-to-proline change at position 47 (mutant TP47; about 10-fold higher). It is believed that other, comparable transposase mutations (in one or more amino acids) that increase binding avidity for OE termini may also be  
10 obtained which would function as well or better in the *in vitro* assay described herein.

One of ordinary skill will also appreciate that changes to the nucleotide sequences of the short repeat sequences of the donor DNA may coordinate with other mutation(s) in or near the  
15 binding region of the transposase enzyme to achieve the same increased binding effect, and the resulting 5- to 50-fold increase in transposition rate. Thus, while the applicants have exemplified one case of a mutation that improves binding of the exemplified transposase, it will be understood that other mutations in the transposase, or in the short repeat sequences, or in both, will also yield transposases that fall within the scope and spirit of the present invention. A suitable method for determining the relative avidity for Tn5 OE termini has been published by Jilk, R. A., et al., "The  
20 Organization of the Outside end of Transposon Tn5," J. Bact.  
25 178:1671-79 (1996).

The transposase of the present invention is also less likely than the wild type protein to assume an inactive multimeric form. In the preferred embodiment, that class (2) mutation from wild type can be achieved by modifying amino acid 372 (leucine) of wild type Tn5 transposase to a proline (and, likewise by modifying the corresponding DNA to encode proline). This mutation, referred to as LP372, has previously been characterized as a mutation in the dimerization region of the transposase. Weinreich, et al., *supra*. It was noted by  
30 Weinreich et al. that this mutation at position 372 maps to a region shown previously to be critical for interaction with an inhibitor of Tn5 transposition. The inhibitor is a protein encoded by the same gene that encodes the transposase, but  
35 which is truncated at the N-terminal end of the protein,  
40 relative to the transposase. The approach of Weinreich et al.

5 for determining the extent to which multimers are formed is suitable for determining whether a mutation falls within the scope of this element.

10 It is thought that when wild type Tn5 transposase multimerizes, its activity *in trans* is reduced. Presumably, a mutation in the dimerization region reduces or prevents multimerization, thereby reducing inhibitory activity and leading to levels of transposition 5- to 50-fold higher than are seen with the wild type transposase. The LP372 mutation achieves about 10-fold higher transposition levels than wild 15 type. Likewise, other mutations (including mutations at a one or more amino acid) that reduce the ability of the transposase to multimerize would also function in the same manner as the single mutation at position 372, and would also be suitable in a transposase of the present invention. It may also be 20 possible to reduce the ability of a Tn5 transposase to multimerize without altering the wild type sequence in the so-called dimerization region, for example by adding into the system another protein or non-protein agent that blocks the dimerization site. Alternatively, the dimerization region 25 could be removed entirely from the transposase protein.

As was noted above, the inhibitor protein, encoded in partially overlapping sequence with the transposase, can interfere with transposase activity. As such, it is desired 30 that the amount of inhibitor protein be reduced over the amount observed in wild type *in vivo*. For the present assay, the transposase is used in purified form, and it may be possible to separate the transposase from the inhibitor (for example, according to differences in size) before use. However, it is 35 also possible to genetically eliminate the possibility of having any contaminating inhibitor protein present by removing its start codon from the gene that encodes the transposase.

An AUG in the wild type Tn5 transposase gene that encodes methionine at transposase amino acid 56 is the first codon of the inhibitor protein. However, it has already been shown that 40 replacement of the methionine at position 56 has no apparent effect upon the transposase activity, but at the same time

5 prevents translation of the inhibitor protein, thus resulting  
in a somewhat higher transposition rate. Weigand, T. W. and W.  
S. Reznikoff, "Characterization of Two Hypertransposing Tn5  
Mutants," J. Bact. 174:1229-1239 (1992), incorporated herein by  
reference. In particular, the present inventors have replaced  
10 the methionine with an alanine in the preferred embodiment (and  
have replaced the methionine-encoding AUG codon with an  
alanine-encoding GCC). A preferred transposase of the present  
invention therefore includes an amino acid other than  
methionine at amino acid position 56, although this change can  
15 be considered merely technically advantageous (since it ensures  
the absence of the inhibitor from the *in vitro* system) and not  
essential to the invention (since other means can be used to  
eliminate the inhibitor protein from the *in vitro* system).

The most preferred transposase amino acid sequence known  
20 to the inventors differs from the wild type at amino acid  
positions 54, 56, and 372. The mutations at positions 54 and  
372 separately contribute approximately a 10-fold increase to  
the rate of transposition reaction *in vivo*. When the mutations  
25 are combined using standard recombinant techniques into a  
single molecule containing both classes of mutations, reaction  
rates of at least about 100-fold higher than can be achieved  
using wild type transposase are observed when the products of  
the *in vitro* system are tested *in vivo*. The mutation at  
30 position 56 does not directly affect the transposase activity.

Other mutants from wild type that are contemplated to be  
likely to contribute to high transposase activity *in vitro*  
include, but are not limited to glutaminic acid-to-lysine at  
position 110, and glutamic acid to lysine at position 345.

It is, of course, understood that other changes apart from  
35 these noted positions can be made to the modified transposase  
(or to a construct encoding the modified transposase) without  
adversely affecting the transposase activity. For example, it  
is well understood that a construct encoding such a transposase  
could include changes in the third position of codons such that  
40 the encoded amino acid does not differ from that described  
herein. In addition, certain codon changes have little or no

5 functional effect upon the transposition activity of the  
encoded protein. Finally, other changes may be introduced  
which provide yet higher transposition activity in the encoded  
protein. It is also specifically envisioned that combinations  
10 of mutations can be combined to encode a modified transposase  
having even higher transposition activity than has been  
exemplified herein. All of these changes are within the scope  
of the present invention. It is noted, however, that a  
modified transposase containing the EK110 and EK345 mutations  
15 (both described by Weigand and Reznikoff, *supra*, had lower  
transposase activity than a transposase containing either  
mutation alone.

After the enzyme is prepared and purified, as described  
*supra*, it can be used in the *in vitro* transposition reaction  
described above to introduce any desired transposable element  
20 from a donor DNA into a target DNA. The donor DNA can be  
circular or can be linear. If the donor DNA is linear, it is  
preferred that the repeat sequences flanking the transposable  
element should not be at the termini of the linear fragment but  
should rather include some DNA upstream and downstream from the  
25 region flanked by the repeat sequences.

As was noted above, Tn5 transposition requires a pair of  
eighteen or nineteen base long termini. The wild type Tn5  
outside end (OE) sequence (5'-CTGACTCTTATACACAAGT-3') (SEQ ID  
NO: 7) has been described. It has been discovered that a  
30 transposase-catalyzed *in vitro* transposition frequency at least  
as high as that of wild type OE is achieved if the termini in a  
construct include bases ATA at positions 10, 11, and 12,  
respectively, as well as the nucleotides in common between wild  
type OE and IE (e.g., at positions 1-3, 5-9, 13, 14, 16, and  
35 optionally 19). The nucleotides at positions 4, 15, 17, and 18  
can correspond to the nucleotides found at those positions in  
either wild type OE or wild type IE. It is noted that the  
transposition frequency can be enhanced over that of wild type  
OE if the nucleotide at position 4 is a T. The importance of  
40 these particular bases to transposition frequency has not  
previously been identified.

5        It is noted that these changes are not intended to  
encompass every desirable modification to OE. As is described  
elsewhere herein, these attributes of acceptable termini  
modifications were identified by screening mutants having  
randomized differences between IE and OE termini. While the  
10      presence in the termini of certain nucleotides is shown herein  
to be advantageous, other desirable terminal sequences may yet  
be obtained by screening a larger array of degenerate mutants  
that include changes at positions other than those tested  
15      herein as well as mutants containing nucleotides not tested in  
the described screening. In addition, it is clear to one  
skilled in the art that if a different transposase is used, it  
may still be possible to select other variant termini, more  
compatible with that particular transposase.

20      Among the mutants shown to be desirable and within the  
scope of the invention are two hyperactive mutant OE sequences  
that were identified *in vivo*. Although presented here as  
single stranded sequences, in fact, the wild type and mutant OE  
sequences include complementary second strands. The first  
25      hyperactive mutant, 5'-CTGTCTCTTATACACATCT-3' (SEQ ID NO: 8),  
differs from the wild type OE sequence at positions 4, 17, and  
18, counting from the 5' end, but retains ATA at positions 10-  
12. The second, 5'-CTGTCTCTTATACAGATCT-3' (SEQ ID NO: 9),  
differs from the wild type OE sequence at positions 4, 15, 17,  
30      and 18, but also retains ATA at positions 10-12. These two  
hyperactive mutant OE sequences differ from one another only at  
position 15, where either G or C is present. OE-like activity  
(or higher activity) is observed in a mutant sequence when it  
contains ATA at positions 10, 11 and 12. It may be possible to  
reduce the length of the OE sequence from 19 to 18 nucleotide  
35      pairs with little or no effect.

When one of the identified hyperactive mutant OE sequences  
flanks a substrate DNA, the *in vivo* transposition frequency of  
EK54/MA56 transposase is increased approximately 40-60 fold  
over the frequency that is observed when wild type OE termini  
40      flank the transposable DNA. The EK54/MA56 transposase is  
already known to have an *in vivo* transposition frequency

5 approximately an 8-10 fold higher than wild type transposase, using wild type OE termini. Tn5 transposase having the EK54/MA56 mutation is known to bind with greater avidity to OE and with lesser avidity to the Tn5 inside ends (IE) than wild type transposase.

10 A suitable mutant terminus in a construct for use in the assays of the present invention is characterized biologically as yielding more papillae per colony in a comparable time, say 68 hours, than is observed in colonies harboring wild type OE in a comparable plasmid. Wild type OE can yield about 100 15 papillae per colony when measured 68 hours after plating in a papillation assay using EK54/MA56 transposase, as is described elsewhere herein. A preferred mutant would yield between about 200 and 3000 papillae per colony, and a more preferred mutant between about 1000 and 3000 papillae per colony, when measured 20 in the same assay and time frame. A most preferred mutant would yield between about 2000 and 3000 papillae per colony when assayed under the same conditions. Papillation levels may be even greater than 3000 per colony, although it is difficult to quantitate at such levels.

25 Transposition frequency is also substantially enhanced in the *in vitro* transposition assay of the present invention when substrate DNA is flanked by a preferred mutant OE sequence and a most preferred mutant transposase (comprising EK54/MA56/LP372 mutations) is used. Under those conditions, essentially all of 30 the substrate DNA is converted into transposition products.

35 The rate of *in vitro* transposition observed using the hyperactive termini is sufficiently high that, in the experience of the inventors, there is no need to select for transformation for further study have shown evidence of transposition events.

40 This advance can represent a significant savings in time and laboratory effort. For example, it is particularly advantageous to be able to improve *in vitro* transposition frequency by modifying DNA rather than by modifying the transposase because as transposase activity increases in host

5   cells, there is an increased likelihood that cells containing  
the transposase are killed during growth as a result of  
aberrant DNA transpositions. In contrast, DNA of interest  
containing the modified OE termini can be grown in sources  
completely separate from the transposase, thus not putting the  
10 host cells at risk.

Without intending to limit the scope of this aspect of  
this invention, it is apparent that the tested hyperactive  
termini do not bind with greater avidity to the transposase  
than do wild type OE termini. Thus, the higher transposition  
15 frequency brought about by the hyperactive termini is not due  
to enhanced binding to transposase.

The transposable element between the termini can include  
any desired nucleotide sequence. The length of the  
transposable element between the termini should be at least  
20 about 50 base pairs, although smaller inserts may work. No  
upper limit to the insert size is known. However, it is known  
that a donor DNA portion of about 300 nucleotides in length can  
function well. By way of non-limiting examples, the  
transposable element can include a coding region that encodes a  
25 detectable or selectable protein, with or without associated  
regulatory elements such as promoter, terminator, or the like.

If the element includes such a detectable or selectable  
coding region without a promoter, it will be possible to  
identify and map promoters in the target DNA that are uncovered  
30 by transposition of the coding region into a position  
downstream thereof, followed by analysis of the nucleic acid  
sequences upstream from the transposition site.

Likewise, the element can include a primer binding site  
that can be transposed into the target DNA, to facilitate  
35 sequencing methods or other methods that rely upon the use of  
primers distributed throughout the target genetic material.  
Similarly, the method can be used to introduce a desired  
restriction enzyme site or polylinker, or a site suitable for  
another type of recombination, such as a cre-lox, into the  
40 target.

The invention can be better understood upon consideration

5 of the following examples which are intended to be exemplary  
and not limiting on the invention.

#### EXAMPLES

To obtain the transposase modified at position 54, the first third of the coding region from an existing DNA clone 10 that encodes the Tn5 transposase but not the inhibitor protein (MA56) was mutagenized according to known methods and DNA fragments containing the mutagenized portion were cloned to produce a library of plasmid clones containing a full length transposase gene. The clones making up the library were 15 transformed into *E. coli* K-12 strain MDW320 bacteria which were plated and grown into colonies. Transposable elements provided in the bacteria on a separate plasmid contained a defective lacZ gene. The separate plasmid, pOXgen386, was described by Weinreich, M. et al., "A functional analysis of the Tn5 20 Transposase: Identification of Domains Required for DNA Binding and Dimerization," J. Mol. Biol. 241:166-177 (1993), incorporated herein by reference. Colonies having elevated 25 transposase activity were selected by screening for blue (LacZ) spots in white colonies grown in the presence of X-gal. This papillation assay was described by Weinreich, et al. (1993), *supra*. The 5'-most third of Tn5 transposase genes from such colonies were sequenced to determine whether a mutation was responsible for the increase in transposase activity. It was determined that a mutation at position 54 to lysine (K) 30 correlated well with the increase in transposase activity. Plasmid pRZ5412-EK54 contains lysine at position 54 as well as the described alanine at position 56.

The fragment containing the LP372 mutation was isolated 35 from pRZ4870 (Weinreich et al (1994)) using restriction enzymes NheI and BglII, and were ligated into NheI-BglII cut pRZ5412-EK54 to form a recombinant gene having the mutations at positions 54, 56 and 372, as described herein and shown in SEQ ID NO:1. The gene was tested and shown to have at least about a one hundred fold increase in activity relative to wild type 40 Tn5 transposase. Each of the mutants at positions 54 and 372

5 alone had about a 10-fold increase in transposase activity.

The modified transposase protein encoded by the triple-mutant recombinant gene was transferred into commercial T7 expression vector pET-21D (commercially available from Novagen, Madison, WI) by inserting a BspHI/SalI fragment into NhoI/XhoI fragment of the pET-21D vector. This cloning puts the modified transposase gene under the control of the T7 promoter, rather than the natural promoter of the transposase gene. The gene product was overproduced in BL21(DE3)pLySS bacterial host cells, which do not contain the binding site for the enzyme, by specific induction in a fermentation process after cell growth is complete. (See, Studier, F. W., et al., "Use of T7 RNA Polymerase to Direct Expression of Cloned Genes," Methods Enzymol. 185:60-89 (1990)). The transposase was partially purified using the method of de la Cruz, modified by inducing overproduction at 33 or 37°C. After purification, the enzyme preparation was stored at -70°C in a storage buffer (10% glycerol, 0.7M NaCl, 20 mM Tris-HCl, pH 7.5, 0.1% Triton-X100 and 10 mM CHAPS) until use. This storage buffer is to be considered exemplary and not optimized.

25 A single plasmid (pRZTL1, Fig. 1) was constructed to serve as both donor and target DNA in this Example. The complete sequence of the pRZTL1 plasmid DNA is shown and described in SEQ ID NO:3. Plasmid pRZTL1 contains two Tn5 19 base pair OE termini in inverted orientation to each other. Immediately adjacent to one OE sequence is a gene that would encode tetracycline resistance, but for the lack of an upstream promoter. However, the gene is expressed if the tetracycline resistance gene is placed downstream of a transcribed region (e.g., under the control of the promoter that promotes transcription of the chloramphenicol resistance gene also present on pRZTL1). Thus, the test plasmid pRZTL1 can be assayed *in vivo* after the *in vitro* reaction to confirm that transposition has occurred. The plasmid pRZTL1 also includes an origin of replication in the transposable element, which ensures that all transposition products are plasmids that can replicate after introduction in host cells.

5       The following components were used in typical 20 $\mu$ l *in vitro* transposition reactions:

Modified transposase: 2  $\mu$ l (approximately 0.1  $\mu$ g enzyme/ $\mu$ l) in storage buffer (10% glycerol, 0.7M NaCl, 20 mM Tris-HCl, pH 7.5, 0.1% Triton-X100 and 10 mM CHAPS)

10       Donor/Target DNA: 18  $\mu$ l (approximately 1-2  $\mu$ g) in reaction buffer (at final reaction concentrations of 0.1 M potassium glutamate, 25 mM Tris acetate, pH 7.5, 10 mM Mg<sup>2+</sup>-acetate, 50  $\mu$ g/ml BSA, 0.5 mM  $\beta$ -mercaptoethanol, 2 mM spermidine, 100  $\mu$ g/ml tRNA).

15       At 20°C, the transposase was combined with pRZTL1 DNA for about 60 minutes. Then, the reaction volume was increased by adding two volumes of reaction buffer and the temperature was raised to 37°C for 2-3 hours whereupon cleavage and strand transfer occurred.

20       Efficient *in vitro* transposition was shown to have occurred by *in vivo* and by *in vitro* methods. *In vivo*, many tetracycline-resistant colonies were observed after transferring the nucleic acid product of the reaction into DH5 $\alpha$  bacterial cells. As noted, tetracycline resistance can only arise in this system if the transposable element is transposed downstream from an active promoter elsewhere on the plasmid. A typical transposition frequency was 0.1% of cells that received plasmid DNA, as determined by counting chloramphenicol resistant colonies. However, this number underestimates the total transposition event frequency because the detection system limits the target to 1/16 of the total.

25       Moreover, *in vitro* electrophoretic (1% agarose) and DNA sequencing analyses of DNA isolated from purified colonies revealed products of true transposition events, including both intramolecular and intermolecular events. Results of typical reactions using circular plasmid pRZTL1 substrates are shown in Lanes 4 & 5. Lane 6 of Fig. 2 shows the results obtained using linear plasmid pRZTL1 substrates.

5       The bands were revealed on 1% agarose gels by staining  
with SYBR Green (FMC BioProducts) and were scanned on a  
Fluorimager SI (Molecular Dynamics). In Figure 2, lane 1 shows  
relaxed circle, linear, and closed circle versions of pRZTL1.  
10      Lanes 2 and 3 show intramolecular and intermolecular  
transposition products after *in vitro* transposition of pRZTL1,  
respectively. The products were purified from electroporated  
DH5 $\alpha$  cells and were proven by size and sequence analysis to be  
genuine transposition products. Lanes 4 and 5 represent  
15      products of two independent *in vitro* reactions using a mixture  
of closed and relaxed circular test plasmid substrates. In  
lane 6, linear pRZTL1 (*Xba*I-cut) was the reaction substrate.  
Lane 7 includes a *Bst*EII digest of lambda DNA as a molecular  
weight standard.

20      Fig. 3 reproduces lanes 4, 5, and 6 of Fig. 2 and shows an  
analysis of various products, based upon secondary restriction  
digest experiments and re-electroporation and DNA sequencing.  
The released donor DNA corresponds to the fragment of pRZTL1  
that contains the kanamycin resistance gene between the two OE  
sequences, or, in the case of the linear substrate, the OE-*Xba*I  
25      fragment. Intermolecular transposition products can be seen  
only as relaxed DNA circles. Intramolecular transposition  
products are seen as a ladder, which results from conversion of  
the initial superhelicity of the substrate into DNA knots. The  
reaction is efficient enough to achieve double transposition  
30      events that are a combination of inter- and intramolecular  
events.

35      A preliminary investigation was made into the nature of  
the termini involved in a transposition reaction. Wild type  
*Tn*5 OE and IE sequences were compared and an effort was  
undertaken to randomize the nucleotides at each of the seven  
positions of difference. A population of oligonucleotides  
degenerate at each position of difference was created. Thus,  
individual oligonucleotides in the population randomly included  
either the nucleotide of the wild type OE or the wild type IE  
40      sequence. In this scheme,  $2^7$  (128) distinct oligonucleotides  
were synthesized using conventional tools. These

5 oligonucleotides having sequence characteristics of both OE and IE are referred to herein as OE/IE-like sequences. To avoid nomenclature issues that arise because the oligonucleotides are intermediate between OE and IE wild type sequences, the applicants herein note that selected oligonucleotide sequences  
10 are compared to the wild type OE rather than to wild type IE, unless specifically noted. It will be appreciated by one skilled in the art that if IE is selected as the reference point, the differences are identical but are identified differently.

15 The following depicts the positions (x) that were varied in this mutant production scheme. WT OE is shown also at SEQ ID NO: 7, WT IE at SEQ ID NO: 10.

20           5' -CTGACTCTTATACACAAGT-3' (WT OE)  
               x       xxx   x xx        (positions of difference)  
          5' -CTGTCTCTTGATCAGATCT-3' (WT IE)

In addition to the degenerate OE/IE-like sequences, the  
37- base long synthetic oligonucleotides also included terminal  
*Sph*I and *Kpn*I restriction enzyme recognition and cleavage sites  
for convenient cloning of the degenerate oligonucleotides into  
25 plasmid vectors. Thus, a library of randomized termini was  
created from population of  $2^7$  (128) types of degenerate  
oligonucleotides.

Fig. 4 shows pRZ1496, the complete sequence of which is  
presented as SEQ ID NO:11. The following features are noted in  
30 the sequence:

<u>Feature</u>	<u>Position</u>
WT OE	94-112
LacZ coding	135-3137
LacY coding	3199-4486
LacA coding	4553-6295
tet <sup>r</sup> coding	6669-9442
transposase coding	10683-12111 (Comp. Strand)
Cassette IE	12184-12225
colE1 sequence	127732-19182

40 The IE cassette shown in Fig. 4 was excised using *Sph*I and *Kpn*I and was replaced, using standard cleavage and ligation methods, by the synthetic termini cassettes comprising OE/IE-

5 like portions. Between the fixed wild type OE sequence and the  
OE/IE-like cloned sequence, plasmid pRZ1496 comprises a gene  
whose activity can be detected, namely LacZYA, as well as a  
selectable marker gene, *tet*<sup>r</sup>. The LacZ gene is defective in  
that it lacks suitable transcription and translation initiation  
10 signals. The LacZ gene is transcribed and translated only when  
it is transposed into a position downstream from such signals.

The resulting clones were transformed using  
electroporation into *dam*<sup>-</sup>, LacZ<sup>-</sup> bacterial cells, in this case  
JCM101/pOXgen cells which were grown at 37°C in LB medium under  
15 standard conditions. A *dam*<sup>-</sup> strain is preferred because *dam*  
methylation can inhibit IE utilization and wild type IE  
sequences include two *dam* methylation sites. A *dam*<sup>-</sup> strain  
eliminates *dam* methylation as a consideration in assessing  
transposition activity. The *Tet*<sup>r</sup> cells selected were LacZ<sup>-</sup>;  
20 transposition-activated Lac expression was readily detectable  
against a negative background. pOXgen is a non-essential F  
factor derivative that need not be provided in the host cells.

In some experiments, the EK54/MA56 transposase was encoded  
directly by the transformed pRZ1496 plasmid. In other  
25 experiments, the pRZ1496 plasmid was modified by deleting a  
unique HindIII/EagI fragment (nucleotides 9112-12083) from the  
plasmid (see Fig. 4) to prevent transposase production. In the  
latter experiments, the host cells were co-transformed with the  
HindIII/EagI-deleted plasmid, termed pRZ5451 (Fig. 4), and with  
30 an EK54/MA56 transposase-encoding chloramphenicol-resistant  
plasmid. In some experiments, a comparable plasmid encoding a  
wild type Tn5 transposase was used for comparison.

Transposition frequency was assessed by a papillation  
assay that measured the number of blue spots (Lac producing  
35 cells or "papillae") in an otherwise white colony. Transformed  
cells were plated (approx. 50 colonies per plate) on Glucose  
minimal Miller medium (Miller, J., Experiments in Molecular  
Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY  
(1972)) containing 0.3% casamino acids, 5-bromo-4-chloro-3-  
40 indolyl-β-D-galactoside (40 µg/ml) and phenyl-β-D-galactoside  
(0.05%). The medium contained tetracycline (15 µg/ml) and,

5 where needed, chloramphenicol (20 µg/ml). Colonies that survived the selection were evaluated for transposition frequency *in vivo*. Although colonies exhibiting superior papillation were readily apparent to the naked eye, the number of blue spots per colony were determined over a period of  
10 several days (approximately 90 hours post-plating).

To show that the high-papillation phenotype was conferred by the end mutations in the plasmids, colonies were re-streaked if they appeared to have papillation levels higher than was observed when wild type IE was included on the plasmid.  
15 Colonies picked from the streaked culture plates were themselves picked and cultured. DNA was obtained and purified from the cultured cells, using standard protocols, and was transformed again into "clean" JCM101/pOXgen cells. Papillation levels were again compared with wild type IE-containing plasmids in the above-noted assays, and consistent results were observed.

To obtain DNA for sequencing of the inserted oligonucleotide, cultures were grown from white portions of 117 hyperpapillating colonies, and DNA was prepared from each 25 colony using standard DNA miniprep methods. The DNA sequence of the OE/IE-like portion of 117 clones was determined (42 from transformations using pRZ1496 as the cloning vehicle; 75 from transformations using pRZ5451 as the cloning vehicle). Only 29 unique mutants were observed. Many mutants were isolated 30 multiple times. All mutants that showed the highest papillation frequencies contain OE-derived bases at positions 10, 11, and 12. When the OE-like bases at these positions were maintained, it was impossible to measure the effect on 35 transposition of other changes, since the papillation level was already extremely high.

One thousand five hundred seventy five colonies were screened as described above. The likelihood that all 128 possible mutant sequences were screened was greater than 95%. Thus, it is unlikely that other termini that contribute to a 40 greater transformation frequency will be obtained using the tested transposase.

TABLE I.  
trans papillation level of hybrid end sequences with EK54 Tnp

mutant	position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	papillation level <sup>a</sup>	# of times isolated <sup>b</sup>
IE	c	t	g	T	c	t	t	G	A	T	c	a	G	a	T	C	t	VL	M	0	0	
OE	A	A	A	A	A	A	A	A	T	A	A	C	C	A	G	A	G	H	H	2	3	
1	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	3	5	
2	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	4	4	
3	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	6	6	
4	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	6	6	
5	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	4	4	
6	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	7	7	
7	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	3	3	
8	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	2	2	
9	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	1	1	
10	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	0	0	
11	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	4	4	
12	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	2	2	
13	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	1	1	
14	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	1	1	
15	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	2	2	
16	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	1	1	
17	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	2	2	
18	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	1	1	
19	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	1	1	
20	A	A	A	A	A	A	A	A	T	A	C	A	A	A	C	A	A	H	H	1	1	
21																						

All hybrid end sequences isolated on PRZ5451 that papillate more frequently than wt IE, when the EK54 Tnp is expressed from pFMA187, are listed. \*trans papillation levels of wt IE, wt OE and hybrid end sequences are classified as follows: VL-very low, L-low, M-medium, and H-high. <sup>b</sup>Although mutants 12 and 13 were not found in this experiment, they were found in cis papillation screening (Table II).

TABLE II.  
cis papillation level of hybrid end sequences with EK54 Tnp

mutant	position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Papillation level <sup>a</sup>	# of times isolated <sup>b</sup>
IE	c	t	g	t	c	t	t	g	a	t	c	a	g	a	t	c	t	L	H	0	2	
OE	A	A	A	A	A	A	A	A	T	A	C	A	C	A	T	C	A	G	H	2	2	
1	A	A	A	A	A	A	A	A	A	A	C	C	A	A	G	A	H	H	H	0	2	
2	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	H	H	H	0	1	
3	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	H	1	1	
4	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	H	1	1	
5	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	H	1	1	
6	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	H	2	2	
7	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	H	3	3	
8	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	H	1	1	
9	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	H	1	1	
10	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	H	1	1	
11	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	H	0	0	
12	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	MH	3	3	
13	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	MH	1	1	
14	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	MH	0	0	
15	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	H	2	2	
16	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	H	1	1	
17	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	M	1	1	
18	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	M	2	2	
19	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	M	2	2	
20	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	M	1	1	
21	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	M	4	4	
22	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	M	1	1	
23	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	M	1	1	
24	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	M	1	1	
25	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	M	1	1	
26	T	T	T	T	T	T	T	T	T	T	C	C	C	C	A	A	G	G	M	1	1	
27	A	A	A	A	A	A	A	A	A	A	C	C	C	C	A	A	G	G	M	2	2	

All hybrid end sequences isolated on PRZ1496 that papillate more frequently than wt IE, when the EK54 Tnp is expressed from the same plasmid, are listed. <sup>a</sup>cis papillation levels of wt IE, wt OE and hybrid end sequences are classified as follows: L-low, M-medium, MH-high, and H-high. <sup>b</sup>Although mutants 2, 10 and 14 were not found in this experiment, they were found in trans papillation screening (Table I).

5       Tables I and II report the qualitative papillation level  
of mutant constructs carrying the indicated hybrid end  
sequences or the wild type OE or IE end sequences. In the  
tables, the sequence at each position of the terminus  
corresponds to wild type IE unless otherwise noted. The  
10      applicants intend that, while the sequences are presented in  
shorthand notation, one of ordinary skill can readily determine  
the complete 19 base pair sequence of every presented mutant,  
and this specification is to be read to include all such  
complete sequences. Table I includes data from trials where  
15      the EK54 transposase was provided *in trans*; Table II, from  
those trials where the EK54 transposase was provided *in cis*.  
Although a transposase provided *in cis* is more active in  
absolute terms than a transposase provided *in trans*, the *cis* or  
trans source of the transposase does not alter the relative *in*  
20      *vivo* transposition frequencies of the tested termini.

Tables I and II show that every mutant that retains ATA at  
positions 10, 11, and 12, respectively, had an activity  
comparable to, or higher than, wild type OE, regardless of  
whether the wild type OE activity was medium (Table I, *trans*)  
25      or high (Table II, *cis*). Moreover, whenever that three-base  
sequence in a mutant was not ATA, the mutant exhibited lower  
papillation activity than wild type OE. It was also noted that  
papillation is at least comparable to, and tends to be  
significantly higher than, wild type OE when position 4 is a T.  
30

Quantitative analysis of papillation levels was difficult,  
beyond the comparative levels shown (very low, low, medium,  
medium high, and high). However, one skilled in the art can  
readily note the papillation level of OE and can recognize  
those colonies having comparable or higher levels. It is  
35      helpful to observe the papillae with magnification.

The number of observed papillae increased over time, as is  
shown in Figs 5 - 7 which roughly quantitate the papillation  
observed in cells transformed separately with 9 clones  
containing either distinct synthetic termini cassettes or wild  
40      type OE or IE termini. In these 3 figures, each mutant is  
identified by its differences from the wild type IE sequence.

5 Note that, among the tested mutants, only mutant 10A/11T/12A had a higher transposition papillation level than wild type OE. That mutant, which would be called mutant 4/15/17/18 when OE is the reference sequence) was the only mutant of those shown in Figs. 5-7 that retained the nucleotides ATA at positions 10,  
10 11, and 12. Figs. 5 (y-axis: 0 - 1500 papillae) and 6 (y-axis:  
0 - 250 papillae) show papillation using various mutants plus IE and OE controls and the EK54/MA56 enzyme. Fig. 7 (y-axis: 0 - 250 papillae), shows papillation when the same mutant sequences were tested against the wild type (more properly, MA56) transposase. The 10A/11T/12A mutant (SEQ ID NO: 9) yielded significantly more papillae (approximately 3000) in a shorter time (68 hours) with ED54/MA56 transposase than was observed even after 90 hours with the WT OE (approximately 1500). A single OE-like nucleotide at position 15 on an IE-like background also increased papillation frequency.  
20

In vivo transposition frequency was also quantitated in a tetracycline-resistance assay using two sequences having high levels of hyperpapillation. These sequences were 5'-  
25 CTGTCTCTTATACACATCT-3' (SEQ ID NO: 8), which differs from the wild type OE sequence at positions 4, 17, and 18, counting from the 5' end, and 5'-CTGTCTCTTATACAGATCT-3' (SEQ ID NO: 9), which differs from the wild type OE at positions 4, 15, 17, and 18. These sequences are considered the preferred mutant termini in an assay using a transposase that contains EK54/MA56 or a  
30 transposase that contains MA56. Each sequence was separately engineered into pRZTL1 in place of the plasmid's two wild type OE sequences. A PCR-amplified fragment containing the desired ends flanking the kanamycin resistance gene was readily cloned into the large HindIII fragment of pRZTL1. The resulting  
35 plasmids are identical to pRZTL1 except at the indicated termini. For comparison, pRZTL1 and a derivative of pRZTL1 containing two wild type IE sequences were also tested. In the assay, JCM101/pOXgen cells were co-transformed with a test plasmid (pRZTL1 or derivative) and a high copy number amp<sup>r</sup>  
40 plasmid that encoded either the EK54/MA56 transposase or wild type (MA56) transposase. The host cells become tetracycline

5           resistant only when a transposition event brings the Tet<sup>r</sup> gene  
into downstream proximity with a suitable transcriptional  
promoter elsewhere on a plasmid or on the chromosome. The  
total number of cells that received the test plasmids was  
determined by counting chloramphenicol resistant, ampicillin  
10          resistant colonies. Transposition frequency was calculated by  
taking the ratio of tet<sup>r</sup>/cam<sup>r</sup>amp<sup>r</sup> colonies. Approximately 40 to  
60 fold increase over wild type OE in *in vivo* transposition was  
observed when using either of the mutant termini and EK54/MA56  
15          transposase. Of the two preferred mutant termini, the one  
containing mutations at three positions relative to the wild  
type OE sequence yielded a higher increase.

As is shown in Fig. 8, which plots the tested plasmid  
against the transposition frequency ( $\times 10^{-8}$ ), little  
transposition was seen when the test plasmid included two IE  
20          termini. Somewhat higher transposition was observed when the  
test plasmid included two OE termini, particularly when the  
EK54/MA56 transposase was employed. In striking contrast, the  
combination of the EK54/MA56 transposase with either of the  
preferred selected ends (containing OE-like bases only at  
25          positions 10, 11, and 12, or positions 10, 11, 12, and 15)  
yielded a great increase in *in vivo* transposition over wild  
type OE termini.

The preferred hyperactive mutant terminus having the most  
30          preferred synthetic terminus sequence 5'-CTGTCTCTTATACACATCT-3'  
(SEQ ID NO: 8) was provided in place of both WT OE termini in  
pRZTL1 (Fig. 4) and was tested in the *in vitro* transposition  
assay of the present invention using the triple mutant  
35          transposase described herein. This mutant terminus was chosen  
for further *in vitro* analysis because its transposition  
frequency was higher than for the second preferred synthetic  
terminus and because it has no dam methylation sites, so dam  
methylation no longer affects transposition frequency. In  
contrast the 4/15/17/18 mutant does have a dam methylation  
site.

40          In a preliminary experiment, CHAPS was eliminated from the  
reaction, but the pre-incubation step was used. The reaction

5 was pre-incubated for 1 hour at 20°C, then diluted two times, and then incubated for 3 hours at 37°C. About 0.5 µg of DNA and 0.4µg of transposase was used. The transposition products were observed on a gel. With the mutant termini, very little of the initial DNA was observed. Numerous bands representing  
10 primary and secondary transposition reaction products were observed. The reaction mixtures were transformed into DH5 $\alpha$  cells and were plated on chloramphenicol-, tetracycline-, or kanamycin-containing plates.

15 Six hundred forty chloramphenicol-resistant colonies were observed. Although these could represent unreacted plasmid, all such colonies tested (n=12) were sensitive to kanamycin, which indicates a loss of donor backbone DNA. All twelve colonies also included plasmids of varied size; 9 of the 12 were characterized as deletion-inversions, the remaining 3 were  
20 simple deletions. Seventy nine tetracycline-resistant colonies were observed, which indicated an activation of the tet $r$  gene by transposition.

25 Eleven kanamycin resistant colonies were observed. This indicated a low percentage of remaining plasmids carrying the donor backbone DNA.

30 In a second, similar test, about 1 µg of plasmid DNA and 0.2 µg transposase were used. In this test, the reaction was incubated without CHAPS at 37°C for 3 hours without pre-incubation or dilution. Some initial DNA was observed in the gel after the 3 hour reaction. After overnight incubation, only transposition products were observed.

35 The 3 hour reaction products were transformed into DH5 $\alpha$  cells and plated as described. About 50% of the chloramphenicol resistant colonies were sensitive to kanamycin and were presumably transposition products.

The invention is not intended to be limited to the foregoing examples, but to encompass all such modifications and variations as come within the scope of the appended claims. It is envisioned that, in addition to the uses specifically noted herein, other applications will be apparent to the skilled molecular biologist. In particular, methods for  
40

5 introducing desired mutations into prokaryotic or eukaryotic DNA are very desirable. For example, at present it is difficult to knock out a functional eukaryotic gene by homologous recombination with an inactive version of the gene that resides on a plasmid. The difficulty arises from the need  
10 to flank the gene on the plasmid with extensive upstream and downstream sequences. Using this system, however, an inactivating transposable element containing a selectable marker gene (e.g., neo) can be introduced *in vitro* into a plasmid that contains the gene that one desires to inactivate.  
15 After transposition, the products can be introduced into suitable host cells. Using standard selection means, one can recover only cell colonies that contain a plasmid having the transposable element. Such plasmids can be screened, for example by restriction analysis, to recover those that contain  
20 a disrupted gene. Such clones can then be introduced directly into eukaryotic cells for homologous recombination and selection using the same marker gene.

Also, one can use the system to readily insert a PCR-amplified DNA fragment into a vector, thus avoiding traditional  
25 cloning steps entirely. This can be accomplished by (1) providing suitable a pair of PCR primers containing OE termini adjacent to the sequence-specific parts of the primers, (2) performing standard PCR amplification of a desired nucleic acid fragment, (3) performing the *in vitro* transposition reaction of  
30 the present invention using the double-stranded products of PCR amplification as the donor DNA.

5

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: Reznikoff, William S  
Gorysin, Igor Y  
Zhou, Hong

10 (ii) TITLE OF INVENTION: System for In Vitro Transposition

(iii) NUMBER OF SEQUENCES: 11

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Quarles & Brady  
(B) STREET: 1 South Pinckney Street  
(C) CITY: Madison  
(D) STATE: WI  
(E) COUNTRY: USA  
(F) ZIP: 53703

20 (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

25 (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:

30 (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Berson, Bennett J  
(B) REGISTRATION NUMBER: 37094  
(C) REFERENCE/DOCKET NUMBER: 960296.94142

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 608/251-5000  
(B) TELEFAX: 608-251-9166

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1534 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Gene encoding modified Tn5  
transposase enzyme"

45 (ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 93..1523

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTGACTCTTA TACACAAGTA GCGTCCTGAA CGGAACCTTT CCCGTTTCC AGGATCTGAT 60

50 CTTCCATGTG ACCTCCTAAC ATGGTAACGT TC ATG ATA ACT TCT GCT CTT CAT 113  
Met Ile Thr Ser Ala Leu His  
1 5

5	CGT GCG GCC GAC TGG GCT AAA TCT GTG TTC TCT TCG GCG GCG CTG GGT Arg Ala Ala Asp Trp Ala Lys Ser Val Phe Ser Ser Ala Ala Leu Gly 10 15 20	161
10	GAT CCT CGC CGT ACT GCC CGC TTG GTT AAC GTC GCC GCC CAA TTG GCA Asp Pro Arg Arg Thr Ala Arg Leu Val Asn Val Ala Ala Gln Leu Ala 25 30 35	209
15	AAA TAT TCT GGT AAA TCA ATA ACC ATC TCA TCA GAG GGT AGT AAA GCC Lys Tyr Ser Gly Lys Ser Ile Thr Ile Ser Ser Glu Gly Ser Lys Ala 40 45 50 55	257
20	GCC CAG GAA GGC GCT TAC CGA TTT ATC CGC AAT CCC AAC GTT TCT GCC Ala Gln Glu Gly Ala Tyr Arg Phe Ile Arg Asn Pro Asn Val Ser Ala 60 65 70	305
25	GAG GCG ATC AGA AAG GCT GGC GCC ATG CAA ACA GTC AAG TTG GCT CAG Glu Ala Ile Arg Lys Ala Gly Ala Met Gln Thr Val Lys Leu Ala Gln 75 80 85	353
30	GAG TTT CCC GAA CTG CTG GCC ATT GAG GAC ACC ACC TCT TTG AGT TAT Glu Phe Pro Glu Leu Leu Ala Ile Glu Asp Thr Thr Ser Leu Ser Tyr 90 95 100	401
35	CGC CAC CAG GTC GCC GAA GAG CTT GGC AAG CTG GGC TCT ATT CAG GAT Arg His Gln Val Ala Glu Leu Gly Lys Leu Gly Ser Ile Gln Asp 105 110 115	449
40	AAA TCC CGC GGA TGG TGG GTT CAC TCC GTT CTC TTG CTC GAG GCC ACC Lys Ser Arg Gly Trp Trp Val His Ser Val Leu Leu Leu Glu Ala Thr 120 125 130 135	497
45	ACA TTC CGC ACC GTA GGA TTA CTG CAT CAG GAG TGG TGG ATG CGC CCG Thr Phe Arg Thr Val Gly Leu Leu His Gln Glu Trp Trp Met Arg Pro 140 145 150	545
50	GAT GAC CCT GCC GAT GCG GAT GAA AAG GAG AGT GGC AAA TGG CTG GCA Asp Asp Pro Ala Asp Ala Asp Glu Lys Glu Ser Gly Lys Trp Leu Ala 155 160 165	593
55	GCG GCC GCA ACT AGC CGG TTA CGC ATG GGC AGC ATG ATG AGC AAC GTG Ala Ala Ala Thr Ser Arg Leu Arg Met Gly Ser Met Met Ser Asn Val 170 175 180	641
60	ATT GCG GTC TGT GAC CGC GAA GCC GAT ATT CAT GCT TAT CTG CAG GAC Ile Ala Val Cys Asp Arg Glu Ala Asp Ile His Ala Tyr Leu Gln Asp 185 190 195	689
65	AGG CTG GCG CAT AAC GAG CGC TTC GTG GTG CGC TCC AAG CAC CCA CGC Arg Leu Ala His Asn Glu Arg Phe Val Val Arg Ser Lys His Pro Arg 200 205 210 215	737
70	AAG GAC GTA GAG TCT GGG TTG TAT CTG ATC GAC CAT CTG AAG AAC CAA Lys Asp Val Glu Ser Gly Leu Tyr Leu Ile Asp His Leu Lys Asn Gln 220 225 230	785
75	CCG GAG TTG GGT GGC TAT CAG ATC AGC ATT CCG CAA AAG GGC GTG GTG Pro Glu Leu Gly Gly Tyr Gln Ile Ser Ile Pro Gln Lys Gly Val Val 235 240 245	833
80	GAT AAA CGC GGT AAA CGT AAA AAT CGA CCA GCC CGC AAG GCG AGC TTG Asp Lys Arg Gly Lys Arg Lys Asn Arg Pro Ala Arg Lys Ala Ser Leu 250 255 260	881
85	AGC CTG CGC AGT GGG CGC ATC ACG CTA AAA CAG GGG AAT ATC ACG CTC Ser Leu Arg Ser Gly Arg Ile Thr Leu Lys Gln Gly Asn Ile Thr Leu 265 270 275	929

5	AAC GCG GTG CTG GCC GAG GAG ATT AAC CCG CCC AAG GGT GAG ACC CCG Asn Ala Val Leu Ala Glu Glu Ile Asn Pro Pro Lys Gly Glu Thr Pro 280 285 290 295	977
10	TTG AAA TGG TTG TTG CTG ACC GGC GAA CCG GTC GAG TCG CTA GCC CAA Leu Lys Trp Leu Leu Thr Gly Glu Pro Val Glu Ser Leu Ala Gln 300 305 310	1025
15	GCC TTG CGC GTC ATC GAC ATT TAT ACC CAT CGC TGG CGG ATC GAG GAG Ala Leu Arg Val Ile Asp Ile Tyr Thr His Arg Trp Arg Ile Glu Glu 315 320 325	1073
20	TTC CAT AAG GCA TGG AAA ACC GGA GCA GGA GCC GAG AGG CAA CGC ATG Phe His Lys Ala Trp Lys Thr Gly Ala Gly Ala Glu Arg Gln Arg Met 330 335 340	1121
25	GAG GAG CCG GAT AAT CTG GAG CGG ATG GTC TCG ATC CTC TCG TTT GTT Glu Glu Pro Asp Asn Leu Glu Arg Met Val Ser Ile Leu Ser Phe Val 345 350 355	1169
30	GCG GTC AGG CTG TTA CAG CTC AGA GAA AGC TTC ACG CCG CCG CAA GCA Ala Val Arg Leu Leu Gln Leu Arg Glu Ser Phe Thr Pro Pro Gln Ala 360 365 370 375	1217
35	CTC AGG GCG CAA GGG CTG CTA AAG GAA GCG GAA CAC GTA GAA AGC CAG Leu Arg Ala Gln Gly Leu Leu Lys Glu Ala Glu His Val Glu Ser Gln 380 385 390	1265
40	TCC GCA GAA ACG GTG CTG ACC CCG GAT GAA TGT CAG CTA CTG GGC TAT Ser Ala Glu Thr Val Leu Thr Pro Asp Glu Cys Gln Leu Leu Gly Tyr 395 400 405	1313
45	CTG GAC AAG GGA AAA CGC AAG CGC AAA GAG AAA GCA GGT AGC TTG CAG Leu Asp Lys Gly Lys Arg Lys Arg Lys Glu Lys Ala Gly Ser Leu Gln 410 415 420	1361
50	TGG GCT TAC ATG GCG ATA GCT AGA CTG GGC GGT TTT ATG GAC AGC AAG Trp Ala Tyr Met Ala Ile Ala Arg Leu Gly Phe Met Asp Ser Lys 425 430 435	1409
55	CGA ACC GGA ATT GCC AGC TGG GGC GCC CTC TGG GAA GGT TGG GAA GCC Arg Thr Gly Ile Ala Ser Trp Gly Ala Leu Trp Glu Gly Trp Glu Ala 440 445 450 455	1457
60	CTG CAA AGT AAA CTG GAT GGC TTT CTT GCC GCC AAG GAT CTG ATG GCG Leu Gln Ser Lys Leu Asp Gly Phe Leu Ala Ala Lys Asp Leu Met Ala 460 465 470	1505
	CAG GGG ATC AAG ATC TGA TCAAGAGACA G Gln Gly Ile Lys Ile * 475	1534

## (2) INFORMATION FOR SEQ ID NO:2:

45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 477 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val			
1	5	10	15





5	GGC ACC GTC ACC CTG GAT GCT GTA GGC ATA GGC TTG GTT ATG CCG GTA Gly Thr Val Thr Leu Asp Ala Val Gly Ile Gly Leu Val Met Pro Val 490 495 500	157
10	CTG CCG GGC CTC TTG CGG GAT ATC GTC CAT TCC GAC AGC ATC GCC AGT Leu Pro Gly Leu Leu Arg Asp Ile Val His Ser Asp Ser Ile Ala Ser 505 510 515 520	205
15	CAC TAT GGC GTG CTG CTA GCG CTA TAT GCG TTG ATG CAA TTT CTA TGC His Tyr Gly Val Leu Leu Ala Leu Tyr Ala Leu Met Gln Phe Leu Cys 525 530 535	253
20	GCA CCC GTT CTC GGA GCA CTG TCC GAC CGC TTT GGC CGC CGC CCA GTC Ala Pro Val Leu Gly Ala Leu Ser Asp Arg Phe Gly Arg Arg Pro Val 540 545 550	301
25	CTG CTC GCT TCG CTA CTT GGA GCC ACT ATC GAC TAC GCG ATC ATG GCG Leu Leu Ala Ser Leu Leu Gly Ala Thr Ile Asp Tyr Ala Ile Met Ala 555 560 565	349
30	ACC ACA CCC GTC CTG TGG ATC CTC TAC GCC GGA CGC ATC GTG GCC GGC Thr Thr Pro Val Leu Trp Ile Leu Tyr Ala Gly Arg Ile Val Ala Gly 570 575 580	397
35	ATC ACC GGC GCC ACA GGT GCG GTT GCT GGC GCC TAT ATC GCC GAC ATC Ile Thr Gly Ala Thr Gly Ala Val Ala Gly Ala Tyr Ile Ala Asp Ile 585 590 595 600	445
40	ACC GAT GGG GAA GAT CGG GCT CGC CAC TTC GGG CTC ATG AGC GCT TGT Thr Asp Gly Glu Asp Arg Ala Arg His Phe Gly Leu Met Ser Ala Cys 605 610 615	493
45	TTC GGC GTG GGT ATG GTG GCA GGC CCC GTG GCC GGG GGA CTG TTG GGC Phe Gly Val Gly Met Val Ala Gly Pro Val Ala Gly Gly Leu Leu Gly 620 625 630	541
50	GCC ATC TCC TTG CAT GCA CCA TTC CTT GCG GCG GCG GTG CTC AAC GGC Ala Ile Ser Leu His Ala Pro Phe Leu Ala Ala Val Leu Asn Gly 635 640 645	589
55	CTC AAC CTA CTA CTG GGC TGC TTC CTA ATG CAG GAG TCG CAT AAG GGA Leu Asn Leu Leu Leu Gly Cys Phe Leu Met Gln Glu Ser His Lys Gly 650 655 660	637
60	GAG CGT CGA CCG ATG CCC TTG AGA GCC TTC AAC CCA GTC AGC TCC TTC Glu Arg Arg Pro Met Pro Leu Arg Ala Phe Asn Pro Val Ser Ser Phe 665 670 675 680	685
65	CGG TGG GCG CGG GGC ATG ACT ATC GTC GCC GCA CTT ATG ACT GTC TTC Arg Trp Ala Arg Gly Met Thr Ile Val Ala Ala Leu Met Thr Val Phe 685 690 695	733
70	TTT ATC ATG CAA CTC GTA GGA CAG GTG CCG GCA GCG CTC TGG GTC ATT Phe Ile Met Gln Leu Val Gly Gln Val Pro Ala Ala Leu Trp Val Ile 700 705 710	781
75	TTC GGC GAG GAC CGC TTT CGC TGG AGC GCG ACG ATG ATC GGC CTG TCG Phe Gly Glu Asp Arg Phe Arg Trp Ser Ala Thr Met Ile Gly Leu Ser 715 720 725	829
80	CTT GCG GTA TTC GGA ATC TTG CAC GCC CTC GCT CAA GCC TTC GTC ACT Leu Ala Val Phe Gly Ile Leu His Ala Leu Ala Gln Ala Phe Val Thr 730 735 740	877
85	GGT CCC GCC ACC AAA CGT TTC GGC GAG AAG CAG GCC ATT ATC GCC GGC Gly Pro Ala Thr Lys Arg Phe Gly Glu Lys Gln Ala Ile Ile Ala Gly 745 750 755 760	925

5	ATG GCG GCC GAC GCG CTG GGC TAC GTC TTG CTG GCG TTC GCG ACG CGA Met Ala Ala Asp Ala Leu Gly Tyr Val Leu Leu Ala Phe Ala Thr Arg 765 770 775	973
10	Gly Trp Met Ala Phe Pro Ile Met Ile Leu Leu Ala Ser Gly Gly Ile 780 785 790	1021
	GGG ATG CCC GCG TTG CAG GCC ATG CTG TCC AGG CAG GTA GAT GAC GAC Gly Met Pro Ala Leu Gln Ala Met Leu Ser Arg Gln Val Asp Asp Asp 795 800 805	1069
15	CAT CAG GGA CAG CTT CAA GGA TCG CTC GCG GCT CTT ACC AGC CTA ACT His Gln Gly Gln Leu Gln Gly Ser Leu Ala Ala Leu Thr Ser Leu Thr 810 815 820	1117
	TCG ATC ACT GGA CCG CTG ATC GTC ACG GCG ATT TAT GCC GCC TCG GCG Ser Ile Thr Gly Pro Leu Ile Val Thr Ala Ile Tyr Ala Ala Ser Ala 825 830 835 840	1165
20	AGC ACA TGG AAC GGG TTG GCA TGG ATT GTA GGC GCC GCC CTA TAC CTT Ser Thr Trp Asn Gly Leu Ala Trp Ile Val Gly Ala Ala Leu Tyr Leu 845 850 855	1213
25	GTC TGC CTC CCC GCG TTG CGT CGC GGT GCA TGG AGC CGG GCC ACC TCG Val Cys Leu Pro Ala Leu Arg Arg Gly Ala Trp Ser Arg Ala Thr Ser 860 865 870	1261
	ACC TGA ATGGAAGCCG GCGGCACCTC GCTAACGGAT TCACCACTCC AAGAATTGGA Thr *	1317
	GCCAATCAAT TCTTGGGGAG AACTGTGAAT GCGCAAACCA ACCCTGGCA GAACATATCC	1377
30	ATCGCGTCCG CCATCTCCAG CAGCCGCACG CGGCCGCATCT CGGGCAGCGT TGGGTCCTGG	1437
	CCACGGGTGC GCATGATCGT GCTCCTGTCG TTGAGGACCC GGCTAGGCTG GCGGGGTTGC	1497
	CTTACTGGTT AGCAGAATGA ATCACCGATA CGCGAGCGAA CGTGAAGCGA CTGCTGCTGC	1557
	AAAACGTCTG CGACCTGAGC AACAAACATGA ATGGTCTTCG GTTTCCGTGT TTCGTAAAGT	1617
	CTGGAAACGC GGAAGTCCCC TACGTGCTGC TGAAGTTGCC CGCAACAGAG AGTGGAACCA	1677
35	ACCGGTGATA CCACGATACT ATGACTGAGA GTCAACGCCA TGAGCGGCCT CATTCTTAT	1737
	TCTGAGTTAC AACAGTCCGC ACCGCTGTCC GGTAGCTCCT TCCGGTGGGC GCGGGGCATG	1797
	ACTATCGTCG CCGCACTTAT GACTGTCTTC TTTATCATGC AACTCGTAGG ACAGGTGCCG	1857
	GCAGCGCCCA ACAGTCCCCC GGCCACGGGG CCTGCCACCA TACCCACGCC GAAACAAGCG	1917
	CCCTGCACCA TTATGTTCCG GATCTGCATC GCAGGATGCT GCTGGCTACC CTGTGGAACA	1977
40	CCTACATCTG TATTAACGAA GCGCTAACCG TTTTATCAG GCTCTGGGAG GCAGAATAAA	2037
	TGATCATATC GTCAATTATT ACCTCCACGG GGAGAGCCTG AGCAAACCTGG CCTCAGGCAT	2097
	TTGAGAAGCA CACGGTCACA CTGCTTCCGG TAGTCAATAA ACCGGTAAAC CAGCAATAGA	2157
	CATAAGCGGC TATTTAACGA CCCTGCCCTG AACCGACGAC CGGGTCGAAT TTGCTTTCGA	2217
	ATTCTGCCA TTCACTCCGCT TATTATCAAT TATTCAGGCG TAGCACCAGG CGTTAAGGG	2277
45	CACCAATAAC TGCCTAAAAA AAATTACGCC CGGCCCTGCC ACTCATCGCA GTACTGTTGT	2337
	AATTCAATTAA GCATTCTGCC GACATGGAAG CCATCACAGA CGGCATGATG AACCTGAATC	2397

5	GCCAGCGGCA TCAGCACCTT GTCGCCTTGC GTATAATATT TGCCCATGGT GAAAACGGGG	2457
	GCGAAGAAGT TGTCCATATT GGCCACGTTT AAATCAAAAC TGGTGAAACT CACCCAGGGA	2517
	TTGGCTGAGA CGAAAAACAT ATTCTCAATA AACCCTTAG GGAAATAGGC CAGGTTTCA	2577
	CCGTAACACG CCACATCTTGC GAATATATG TGTAGAAAATC GCCGGAAATC GTCGTGGTAT	2637
	TCACTCCAGA GCGATGAAAA CGTTTCAGTT TGCTCATGGA AAACGGTGTAA ACAAGGGTGA	2697
10	ACACTATCCC ATATCACCAAG CTCACCGTCT TTCATTGCCA TACGGAATTC CGGATGAGCA	2757
	TTCATCAGGC GGGCAAGAAT GTGAATAAAG GCCGGATAAA ACTTGTGCTT ATTTTCTTT	2817
	ACGGTCTTTA AAAAGGCCGT AATATCCAGC TGAACGGTCT GGTTATAGGT ACATTGAGCA	2877
	ACTGACTGAA ATGCCTCAAA ATGTTCTTTA CGATGCCATT GGGATATATC AACGGTGGTA	2937
	TATCCAGTGA TTTTTTCTC CATTTCAGCT TCCTTAGCTC CTGAAAATCT CGATAACTCA	2997
15	AAAAATACGC CCGGTAGTGA TCTTATTCA TTATGGTGA AGTTGGAACC TCTTACGTGC	3057
	CGATCAACGT CTCATTTCG CCAAAAGTTG GCCCAGGGCT TCCCGGTATC AACAGGGACA	3117
	CCAGGATTAA TTTATTCTGC GAAGTGATCT TCCGTCACAG GTATTTATTC GGCGCAAAGT	3177
	GCGTCGGGTG ATGCTGCCAA CTTACTGATT TAGTGTATGA TGGTGTGTTT GAGGTGCTCC	3237
	AGTGGTTCT GTTTCTATCA GCTGTCCCTC CTGTTCAGCT ACTGACGGGG TGGTGCCTAA	3297
20	CGGAAAAGC ACCGCCGGAC ATCAGCGCTA GCGGAGTGTAA TACTGGCTTA CTATGTTGGC	3357
	ACTGATGAGG GTGTCAGTGA AGTGCTTCAT GTGGCAGGAG AAAAAAGGCT GCACCGGTGC	3417
	GTCAGCAGAA TATGTGATAC AGGATATATT CCGCTTCCTC GCTCACTGAC TCGCTACGCT	3477
	CGGTCGTTCG ACTGCCCGA GCGGAAATGG CTTACGAACG GGGCGGAGAT TTCCCTGGAAG	3537
	ATGCCAGGAA GATACTAAC AGGGAAAGTGA GAGGGCCGCG GCAAAGCCGT TTTTCCATAG	3597
25	GCTCCGCCCC CCTGACAAGC ATCACGAAAT CTGACGCTCA AATCAGTGGT GGC GAAACCC	3657
	GACAGGACTA TAAAGATACC AGGCCTTCC CCTGGCGGCT CCCTCGTGC CTCTCCTGTT	3717
	CCTGCCTTTC GGTTTACCGG TGTCAATTCCG CTGTTATGGC CGCGTTGTC TCATTCCACG	3777
	CCTGACACTC AGTTCCGGGT AGGCAGTTCG CTCCAAGCTG GACTGTATGC ACGAACCCCC	3837
	CGTTCACTCC GACCGCTGCG CCTTATCCGG TAACTATCGT CTTGAGTCCA ACCCGGAAAG	3897
30	ACATGCAAAA GCACCACTGG CAGCAGCCAC TGGTAATTGA TTTAGAGGAG TTAGTCTTGA	3957
	AGTCATGCGC CGGTTAAGGC TAAACTGAAA GGACAAGTTT TGGTGACTGC GCTCCTCCAA	4017
	GCCAGTTACC TCGGTTCAAA GAGTTGGTAG CTCAGAGAAC CTTCGAAAAA CCGCCCTGCA	4077
	AGGCGGTTTT TTCGTTTCA GAGCAAGAGA TTACGCGCAG ACCAAAACGA TCTCAAGAAG	4137
	ATCATCTTAT TAATCAGATA AAATATTCT AGAGGTGAAC CATCACCCCTA ATCAAGTTTT	4197
35	TTGGGGTCGA GGTGCCGTAA AGCACTAAAT CGGAACCCCTA AAGGGATGCC CCGATTAGA	4257
	GCTTGACGGG GAAAGCCGGC GAACGTGGCG AGAAAGGAAG GGAAGAAAGC GAAAGGAGCG	4317
	GGCGCTAGGG CGCTGGCAAG TGTAGCGGTC ACGCTGCGCG TAACCACCAC ACCCGCCGCG	4377
	CTTAATGCCTC CGCTACAGCG CCATTGCCA TTCAGGCTGC GCAACTGTTG GGAAGGGCGA	4437

5	TCGGTGCAGG CCTCTTCGCT ATTACGCCAG CTGGCGAAAG GGGGATGTGC TGCAAGGCGA TTAACGTTGGG TAACGCCAGG GTTTCCAG TCACGACGTT GTAAAACGAC GGCCAGTGCC AAGCTTACTT GTGTATAAGA GTCAGTCGAC CTGCAGGGGG GGGGGGGAAA GCCACGTTGT GTCTCAAAAT CTCTGATGTT ACATTGCACA AGATAAAAAT ATATCATCAT GAACAATAAA	4497 4557 4617 4677
10	ACTGTCTGCT TACATAAAC A GTAATACAAG GGGTGTT ATG AGC CAT ATT CAA CGG Met Ser His Ile Gln Arg 1 5	4732
	GAA ACG TCT TGC TCG AGG CCG CGA TTA AAT TCC AAC ATG GAT GCT GAT Glu Thr Ser Cys Ser Arg Pro Arg Leu Asn Ser Asn Met Asp Ala Asp 10 15 20	4780
15	TTA TAT GGG TAT AAA TGG GCT CGC GAT AAT GTC GGG CAA TCA GGT GCG Leu Tyr Gly Tyr Lys Trp Ala Arg Asp Asn Val Gly Gln Ser Gly Ala 25 30 35	4828
20	ACA ATC TAT CGA TTG TAT GGG AAG CCC GAT GCG CCA GAG TTG TTT CTG Thr Ile Tyr Arg Leu Tyr Gly Lys Pro Asp Ala Pro Glu Leu Phe Leu 40 45 50	4876
	AAA CAT GGC AAA GGT AGC GTT GCC AAT GAT GTT ACA GAT GAG ATG GTC Lys His Gly Lys Gly Ser Val Ala Asn Asp Val Thr Asp Glu Met Val 55 60 65 70	4924
25	AGA CTA AAC TGG CTG ACG GAA TTT ATG CCT CTT CCG ACC ATC AAG CAT Arg Leu Asn Trp Leu Thr Glu Phe Met Pro Leu Pro Thr Ile Lys His 75 80 85	4972
	TTT ATC CGT ACT CCT GAT GCA TGG TTA CTC ACC ACT GCG ATC CCC Phe Ile Arg Thr Pro Asp Asp Ala Trp Leu Leu Thr Thr Ala Ile Pro 90 95 100	5020
30	GGG AAA ACA GCA TTC CAG GTA TTA GAA GAA TAT CCT GAT TCA GGT GAA Gly Lys Thr Ala Phe Gln Val Leu Glu Glu Tyr Pro Asp Ser Gly Glu 105 110 115	5068
35	AAT ATT GTT GAT GCG CTG GCA GTG TTC CTG CGC CGG TTG CAT TCG ATT Asn Ile Val Asp Ala Leu Ala Val Phe Leu Arg Arg Leu His Ser Ile 120 125 130	5116
	CCT GTT TGT AAT TGT CCT TTT AAC AGC GAT CGC GTA TTT CGT CTC GCT Pro Val Cys Asn Cys Pro Phe Asn Ser Asp Arg Val Phe Arg Leu Ala 135 140 145 150	5164
40	CAG GCG CAA TCA CGA ATG AAT AAC GGT TTG GTT GAT GCG AGT GAT TTT Gln Ala Gln Ser Arg Met Asn Asn Gly Leu Val Asp Ala Ser Asp Phe 155 160 165	5212
	GAT GAC GAG CGT AAT GGC TGG CCT GTT GAA CAA GTC TGG AAA GAA ATG Asp Asp Glu Arg Asn Gly Trp Pro Val Glu Gln Val Trp Lys Glu Met 170 175 180	5260
45	CAT AAG CTT TTG CCA TTC TCA CCG GAT TCA GTC GTC ACT CAT GGT GAT His Lys Leu Leu Pro Phe Ser Pro Asp Ser Val Val Thr His Gly Asp 185 190 195	5308
50	TTC TCA CTT GAT AAC CTT ATT TTT GAC GAG GGG AAA TTA ATA GGT TGT Phe Ser Leu Asp Asn Leu Ile Phe Asp Glu Gly Lys Leu Ile Gly Cys 200 205 210	5356
	ATT GAT GTT GGA CGA GTC GGA ATC GCA GAC CGA TAC CAG GAT CTT GCC Ile Asp Val Gly Arg Val Gly Ile Ala Asp Arg Tyr Gln Asp Leu Ala 215 220 225 230	5404

5	ATC CTA TGG AAC TGC CTC GGT GAG TTT TCT CCT TCA TTA CAG AAA CGG Ile Leu Trp Asn Cys Leu Gly Glu Phe Ser Pro Ser Leu Gln Lys Arg 235 240 245	5452
10	CTT TTT CAA AAA TAT GGT ATT GAT AAT CCT GAT ATG AAT AAA TTG CAG Leu Phe Gln Lys Tyr Gly Ile Asp Asn Pro Asp Met Asn Lys Leu Gln 250 255 260	5500
	TTT CAT TTG ATG CTC GAT GAG TTT TTC TAA TCAGAATTGG TTAATTGGTT Phe His Leu Met Leu Asp Glu Phe Phe * 265 270	5550
15	GTAACACTGG CAGAGCATTA CGCTGACTTG ACGGGACGGC GGCTTGTTG AATAAATCGA ACTTTGCTG AGTTGAAGGA TCAGATCACG CATCTTCCCG ACAACGCAGA CCGTTCCGTG GCAGAGAAA AGTTCAAAAT CACCAACTGG TCCACCTACA ACAGAGCTCT CATCAACCCT GGCTCCCTCA CTTTCTGGCT GGATGATGGG GCGATTCAAGG CCTGGTATGA GTCAGCAACA CCTTCTTCAC GAGGCAGACC TCAGCGCCCC CCCCCCCCTG CAGGTCGA	5610 5670 5730 5790 5838
(2) INFORMATION FOR SEQ ID NO:4:		
20	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 397 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: protein	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:	
	Met Lys Ser Asn Asn Ala Leu Ile Val Ile Leu Gly Thr Val Thr Leu 1 5 10 15	
	Asp Ala Val Gly Ile Gly Leu Val Met Pro Val Leu Pro Gly Leu Leu 20 25 30	
30	Arg Asp Ile Val His Ser Asp Ser Ile Ala Ser His Tyr Gly Val Leu 35 40 45	
	Leu Ala Leu Tyr Ala Leu Met Gln Phe Leu Cys Ala Pro Val Leu Gly 50 55 60	
35	Ala Leu Ser Asp Arg Phe Gly Arg Arg Pro Val Leu Leu Ala Ser Leu 65 70 75 80	
	Leu Gly Ala Thr Ile Asp Tyr Ala Ile Met Ala Thr Thr Pro Val Leu 85 90 95	
	Trp Ile Leu Tyr Ala Gly Arg Ile Val Ala Gly Ile Thr Gly Ala Thr 100 105 110	
40	Gly Ala Val Ala Gly Ala Tyr Ile Ala Asp Ile Thr Asp Gly Glu Asp 115 120 125	
	Arg Ala Arg His Phe Gly Leu Met Ser Ala Cys Phe Gly Val Gly Met 130 135 140	
45	Val Ala Gly Pro Val Ala Gly Gly Leu Leu Gly Ala Ile Ser Leu His 145 150 155 160	
	Ala Pro Phe Leu Ala Ala Val Leu Asn Gly Leu Asn Leu Leu Leu 165 170 175	

5	Gly Cys Phe Leu Met Gln Glu Ser His Lys Gly Glu Arg Arg Pro Met 180 185 190
	Pro Leu Arg Ala Phe Asn Pro Val Ser Ser Phe Arg Trp Ala Arg Gly 195 200 205
10	Met Thr Ile Val Ala Ala Leu Met Thr Val Phe Phe Ile Met Gln Leu 210 215 220
	Val Gly Gln Val Pro Ala Ala Leu Trp Val Ile Phe Gly Glu Asp Arg 225 230 235 240
	Phe Arg Trp Ser Ala Thr Met Ile Gly Leu Ser Leu Ala Val Phe Gly 245 250 255
15	Ile Leu His Ala Leu Ala Gln Ala Phe Val Thr Gly Pro Ala Thr Lys 260 265 270
	Arg Phe Gly Glu Lys Gln Ala Ile Ile Ala Gly Met Ala Ala Asp Ala 275 280 285
20	Leu Gly Tyr Val Leu Leu Ala Phe Ala Thr Arg Gly Trp Met Ala Phe 290 295 300
	Pro Ile Met Ile Leu Leu Ala Ser Gly Gly Ile Gly Met Pro Ala Leu 305 310 315 320
	Gln Ala Met Leu Ser Arg Gln Val Asp Asp Asp His Gln Gly Gln Leu 325 330 335
25	Gln Gly Ser Leu Ala Ala Leu Thr Ser Leu Thr Ser Ile Thr Gly Pro 340 345 350
	Leu Ile Val Thr Ala Ile Tyr Ala Ala Ser Ala Ser Thr Trp Asn Gly 355 360 365
30	Leu Ala Trp Ile Val Gly Ala Ala Leu Tyr Leu Val Cys Leu Pro Ala 370 375 380
	Leu Arg Arg Gly Ala Trp Ser Arg Ala Thr Ser Thr * 385 390 395

(2) INFORMATION FOR SEQ ID NO:5:



5	Val Asp Ala Ser Asp Phe Asp Asp Glu Arg Asn Gly Trp Pro Val Glu 165 170 175
	Gln Val Trp Lys Glu Met His Lys Leu Leu Pro Phe Ser Pro Asp Ser 180 185 190
10	Val Val Thr His Gly Asp Phe Ser Leu Asp Asn Leu Ile Phe Asp Glu 195 200 205
	Gly Lys Leu Ile Gly Cys Ile Asp Val Gly Arg Val Gly Ile Ala Asp 210 215 220
	Arg Tyr Gln Asp Leu Ala Ile Leu Trp Asn Cys Leu Gly Glu Phe Ser 225 230 235 240
15	Pro Ser Leu Gln Lys Arg Leu Phe Gln Lys Tyr Gly Ile Asp Asn Pro 245 250 255
	Asp Met Asn Lys Leu Gln Phe His Leu Met Leu Asp Glu Phe Phe * 260 265 270

(2) INFORMATION FOR SEQ ID NO:7:

- 20           (i) SEQUENCE CHARACTERISTICS:  
               (A) LENGTH: 19 base pairs  
               (B) TYPE: nucleic acid  
               (C) STRANDEDNESS: double  
               (D) TOPOLOGY: linear

25           (ii) MOLECULE TYPE: other nucleic acid  
               (A) DESCRIPTION: /desc = "Tn5 wild type outside end"  
  
               (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CTGACTCTTA TACACAAAGT

19

(2) INFORMATION FOR SEQ ID NO:8:

- 30           (i) SEQUENCE CHARACTERISTICS:  
               (A) LENGTH: 19 base pairs  
               (B) TYPE: nucleic acid  
               (C) STRANDEDNESS: double  
               (D) TOPOLOGY: linear

35           (ii) MOLECULE TYPE: other nucleic acid  
               (A) DESCRIPTION: /desc = "Tn5 mutant outside end"  
  
               (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CTGTCTCTTA TACACATCT

19

(2) INFORMATION FOR SEQ ID NO:9:

- 40           (i) SEQUENCE CHARACTERISTICS:  
               (A) LENGTH: 19 base pairs  
               (B) TYPE: nucleic acid  
               (C) STRANDEDNESS: double  
               (D) TOPOLOGY: linear

45           (ii) MOLECULE TYPE: other nucleic acid  
               (A) DESCRIPTION: /desc = "Tn5 mutant outside end"  
  
               (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CTGTCTCTTA TACAGATCT

19

## 5 (2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 19 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: other nucleic acid  
 (A) DESCRIPTION: /desc = "Tn5 wild type inside end"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTGTCTCTTG ATCAGATCT

19

## 15 (2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 19182 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: circular

20 (ii) MOLECULE TYPE: other nucleic acid  
 (A) DESCRIPTION: /desc = "Plasmid pRZ4196"

25 (ix) FEATURE:  
 (A) NAME/KEY: repeat\_unit  
 (B) LOCATION: 94..112  
 (D) OTHER INFORMATION: /note= "Wild type OE sequence"

30 (ix) FEATURE:  
 (A) NAME/KEY: repeat\_unit  
 (B) LOCATION: 12184..12225  
 (D) OTHER INFORMATION: /note= "Cassette IE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TTCCCTGTAAC AATAGCAATA CCCCAAATAC CTAATGTAGT TCCAGCAAGC AAGCTAAAAA	60
GTAAAGCAAC AACATAACTC ACCCCTGCAT CTGCTGACTC TTATACACAA GTAGCGTCCC	120
GGGATCGGGA TCCCCTCGTT TTACAACGTC GTGACTGGGA AAACCCCTGGC GTTACCCAAC	180
35 TTAATCGCCT TGCAGCACAT CCCCCCTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA	240
CCGATCGCCC TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGCGCTTT GCCTGGTTTC	300
CGGCACCAGA AGCGGTGCCG GAAAGCTGGC TGGAGTGCAG TCTTCCTGAG GCCGATACTG	360
TCGTCGTCCC CTCAAAATGG CAGATGCACG GTTACGATGC GCCCATCTAC ACCAACGTAA	420
CCTATCCCAT TACGGTCAAT CCGCCGTTG TTCCCACGGA GAATCCGACG GGTTGTTACT	480
40 CGCTCACATT TAATGTTGAT GAAAGCTGGC TACAGGAAGG CCAGACGCGA ATTATTTTG	540
ATGGCGTTAA CTCGGCGTTT CATCTGTGGT GCAACGGGCG CTGGGTCGGT TACGGCCAGG	600
ACAGTCGTTT GCCGTCTGAA TTTGACCTGA GCGCATTTC ACAGCGCCGG AAAAAACGCC	660
TCGCGGTGAT GGTGCTGCGT TGGAGTGACG GCAGTTATCT GGAAGATCAG GATATGTGGC	720
GGATGAGCGG CATTTCCTCGT GACGTCTCGT TGCTGCATAA ACCGACTACA CAAATCAGCG	780
45 ATTTCCATGT TGCCACTCGC TTTAATGATG ATTTCAAGCCG CGCTGTACTG GAGGCTGAAG	840
TTCAGATGTG CGGCGAGTTG CGTGAATACC TACGGGTAAC AGTTCTTTA TGGCAGGGTG	900

5	AAACGCAGGT CGCCAGCGC ACCGCGCCTT TCGGCGGTGA AATTATCGAT GAGCGTGGTG GTTATGCCGA TCGCGTCACA CTACGTCTGA ACGTCGAAAA CCCGAAACTG TGGAGCGCCG AAATCCGAA TCTCTATCGT GCGGTGGTTG AACTGCACAC CGCCGACGGC ACGCTGATTG AAGCAGAACG CTGCGATGTC GGTTCCGCG AGGTGCGGAT TGAAAATGGT CTGCTGCTGC TGAACGGCAA GCCGTTGCTG ATTTCGAGGCG TTAACCGTCA CGAGCATCAT CCTCTGCATG GTCAGGTCAT GGATGAGCAG ACGATGGTGC AGGATATCCT GCTGATGAAG CAGAACAACT TTAACGCCGT GCGCTGTTCG CATTATCCGA ACCATCCGCT GTGGTACACG CTGTGGGACC GCTACGGCCT GTATGTGGTG GATGAAGCCA ATATTGAAAC CCACGGCATG GTGCCAATGA ATCGTCTGAC CGATGATCCG CGCTGGCTAC CGGCGATGAG CGAACCGCTA ACGCGAATGG TGCAGCGCGA TCGTAATCAC CCGAGTGTGA TCATCTGGTC GCTGGGAAT GAATCAGGCC 15 ACGGCGCTAA TCACGACGGC CTGTATCGCT GGATCAAATC TGTCGATCCT TCCCGCCCGG TGCAGTATGA AGGCAGCGGA GCCGACACCA CGGCCACCGA TATTATTTGC CCGATGTACG CGCGCGTGGGA TGAAGACCAG CCCTTCCCGG CTGTGCCGAA ATGGTCCATC AAAAATGGC TTTCGCTACC TGGAGAGACG CGCCCGCTGA TCCTTTGCGA ATACGCCAC GCGATGGGTA ACAGTCTTGG CGGTTTCGCT AAATACTGGC AGGCAGTTCG TCAGTATCCC CGTTTACAGG 20 GCGGCTTCGT CTGGGACTGG GTGGATCAGT CGCTGATTAA ATATGATGAA AACGGCAACC CGTGGTCGGC TTACGGCGGT GATTTGGCG ATACGCCGAA CGATGCCAG TTCTGTATGA ACGGTCTGGT CTTTGCCGAC CGCACGCCGC ATCCAGCGCT GACGGAAGCA AAACACCAGC AGCAGTTTT CCAGTTCCGT TTATCCGGGC AAACCATCGA AGTGACCAGC GAATACCTGT TCCGTCTAG CGATAACGAG CTCCTGCACT GGATGGTGGC GCTGGATGGT AAGCCGCTGG 25 CAAGCGGTGA AGTGCCTCTG GATTCGCTC CACAAGGTAA ACAGTTGATT GAACTGCCTG AACTACCGCA GCCGGAGAGC GCCGGGCAAC TCTGGCTCAC AGTACCGCTA GTGCAACCGA ACGCGACCGC ATGGTCAGAA GCCGGGCACA TCAGCGCCTG GCAGCAGTGG CGTCTGGCGG AAAACCTCAG TGTGACGCTC CCCGCCCGT CCCACGCCAT CCCGCATCTG ACCACCAGCG AAATGGATT TTGCATCGAG CTGGGATAATA AGCGTTGGCA ATTTAACCGC CAGTCAGGCT 30 TTCTTCACA GATGTGGATT GGCGATAAAA ACAACTGCT GACGCCGCTG CGCGATCAGT TCACCCGTGC ACCGCTGGAT AACGACATTG GCGTAAGTGA AGCGACCCGC ATTGACCCTA ACGCCTGGGT CGAACGCTGG AAGGCCGGCGG GCCATTACCA GGCGGAAGCA GCGTTGTTGC AGTGCACGGC AGATAACACTT GCTGATGCGG TGCTGATTAC GACCGCTCAC GCGTGGCAGC ATCAGGGAA AACCTTATTT ATCAGCCGGA AAACCTACCG GATTGATGGT AGTGGTCAAA 35 TGGCGATTAC CGTTGATGTT GAAGTGGCGA GCGATAACCC GCATCCGGCG CGGATTGGCC TGAACTGCCA GCTGGCGCAG GTAGCAGAGC GGGTAAACTG GCTCGGATTA GGGCCGCAAG AAAACATATCC CGACCGCCTT ACTGCCGCCT GTTTGACCG CTGGGATCTG CCATTGTCAG ACATGTATAAC CCCGTACGTC TTCCCGAGCG AAAACGGTCT GCGCTGCCGG ACGCGCGAAT
---	---

5	TGAATTATGG CCCACACCAG TGGCGCGCG ACTTCCAGTT CAACATCAGC CGCTACAGTC AACAGCAACT GATGGAAACC AGCCATCGCC ATCTGCTGCA CGCGGAAGAA GGCACATGGC TGAATATCGA CGGTTCCAT ATGGGGATTG GTGGCGACGA CTCCTGGAGC CCGTCAGTAT CGGCGGATTC CAGCTGAGCG CCGGTCGCTA CCATTACCAAG TTGGTCTGGT GTCAAAAATA ATAATAACCG GGCAGGCCAT GTCTGCCGT ATTCGCGTA AGGAAATCCA TTATGTACTA	3000 3060 3120 3180 3240
10	TTTAAAAAAC ACAAACTTT GGATGTTCGG TTTATTCTT TTCTTTACT TTTTTATCAT GGGAGCCTAC TTCCCCTTT TCCCGATTTG GCTACATGAC ATCAACCATA TCAGCAAAAG TGATAACGGGT ATTATTTTG CCGCTATTC TCTGTTCTCG CTATTATTCC AACCGCTGTT TGGTCTGCTT TCTGACAAAC TCGGGCTGCG CAAATACCTG CTGTGGATTA TTACCGGCAT GTTAGTGATG TTTGCCGT TCTTTATTT TATCTTCGGG CCACTGTTAC AATACAACAT	3300 3360 3420 3480 3540
15	TTTAGTAGGA TCGATTGTTG GTGGTATTAA TCTAGGCTTT TGTTTAACG CCGGTGCGCC AGCAGTAGAG GCATTTATTG AGAAAGTCAG CCGTCGCAGT AATTCGAAT TTGGTGCAGC GCGGATGTTT GGCTGTGTTG GCTGGCGCT GTGTGCCTCG ATTGTCGGCA TCATGTTCAC CATCAATAAT CAGTTGTTT TCTGGCTGGG CTCTGGCTGT GCACTCATCC TCGCCGTTTT ACTCTTTTC GCCAAAACGG ATGCGCCCTC TTCTGCCACG GTTCCAATG CGGTAGGTGC	3600 3660 3720 3780 3840
20	CAACCATTG GCATTTAGCC TTAAGCTGGC ACTGGAACGT TTCAGACAGC CAAAACGTG GTTTTGTCA CTGTATGTTA TTGGCGTTTC CTGCACCTAC GATGTTTTG ACCAACAGTT TGCTAATTTC TTTACTTCGT TCTTGCTAC CGGTGAACAG GGTACGCGGG TATTGGCTA CGTAACGACA ATGGGCGAAT TACTAACGC CTCGATTATG TTCTTGCGC CACTGATCAT TAATCGCATH GGTGGGAAAA ACGCCCTGCT GCTGGCTGGC ACTATTATGT CTGTACGTAT	3900 3960 4020 4080 4140
25	TATTGGCTCA TCGTTGCCA CCTCAGCGCT GGAAGTGGTT ATTCTGAAAA CGCTGCATAT GTTTGAAGTA CCGTTCCCTGC TGGTGGCTG CTTTAAATAT ATTACCAAGCC AGTTGAAGT GCGTTTTCA GCGACGATTG ATCTGGTCTG TTTCTGCTTC TTTAAGCAAC TGGCGATGAT TTTTATGTCT GTACTGGCGG GCAATATGTA TGAAAGCATC GGTTCCAGG GCGCTTATCT GGTGTGGGT CTGGTGGCGC TGGGCTTCAC CTTAATTTC GTGTTCACGC TTAGCGGCC	4200 4260 4320 4380 4440
30	CGGCCCCGCTT TCCCTGCTGC GTCGTCAGGT GAATGAAGTC GCTTAAGCAA TCAATGTCGG ATGCGCGCG ACGCTTATCC GACCAACATA TCATAACGGA GTGATCGCAT TGAACATGCC AATGACCGAA AGAATAAGAG CAGGCAAGCT ATTTACCGAT ATGTGCGAAG GCTTACCGGA AAAAAGACTT CGTGGGAAAA CGTTAATGTA TGAGTTAAT CACTCGCATC CATCAGAAGT TGAAAAAAAGA GAAAGCCTGA TTAAAGAAAT GTTGCCACG GTAGGGAAA ACGCCTGGGT	4500 4560 4620 4680 4740
35	AGAACCGCCT GTCTATTTC CTTACGGTTC CAACATCCAT ATAGGCCGA ATTTTTATGC AAATTCAAT TTAACCATTG TCGATGACTA CACGGTAACA ATCGGTGATA ACGTACTGAT TGCACCCAAAC GTTACTCTTT CGGTTACGGG ACACCCCTGTA CACCATGAAT TGAGAAAAAA CGGCGAGATG TACTCTTTTC CGATAACGAT TGGCAATAAC GTCTGGATCG GAAGTCATGT	4800 4860 4920 4980

5	GGTTATTAAT CCAGGCGTCA CCATCGGGGA TAATTCTGTT ATTGGCGCGG GTAGTATCGT CACAAAAGAC ATTCCACCAA ACGTCGTGGC GGCTGGCGTT CCTTGTGCGG TTATTCGCGA AATAAACGAC CGGGATAAGC ACTATTATTT CAAAGATTAT AAAGTTGAAT CGTCAGTTA AATTATAAAA ATTGCCTGAT ACGCTGCGCT TATCAGGCC ACAAGTTCAAG CGATCTACAT TAGCCGCATC CGGCATGAAC AAAGCGCAGG AACAAAGCGTC GCATCATGCC TCTTGACCC	5040 5100 5160 5220 5280
10	ACAGCTGCGG AAAACGTACT GGTGCAAAAC GCAGGGTTAT GATCATCAGC CCAACGACGC ACAGCGCATG AAATGCCAG TCCATCAGGT AATTGCCGCT GATACTACGC AGCACGCCAG AAAACCACGG GGCAAGCCCG GCGATGATAA AACCGATTCC CTGCATAAAC GCCACCAGCT TGCCAGCAAT AGCCGGTTGC ACAGAGTGAT CGAGCGCCAG CAGCAAACAG AGCGGAAACG CGCCGCCAG ACCTAACCCA CACACCATCG CCCACAATAC CGGCAATTGC ATCGGCAGCC	5340 5400 5460 5520 5580
15	AGATAAAAGCC GCAGAACCCCC ACCAGTTGTA ACACCAGCGC CAGCATTAAAC AGTTTGCGCC GATCCTGATG GCGAGCCATA GCAGGCATCA GCAAAGCTCC TGCGGCTTGC CCAAGCGTCA TCAATGCCAG TAAGGAACCG CTGTACTGCG CGCTGGCACC AATCTCAATA TAGAAAGCGG GTAACCAGGC AATCAGGCTG GCGTAACCAGC CGTTAATCAG ACCGAAGTAA ACACCCAGCG TCCACGGCGCG GGGAGTGAAT ACCACCGCAA CCGGAGTGGT TGTGTCTTG TGGGAAGAGG	5640 5700 5760 5820 5880
20	CGACCTCGCG GGCGCTTGC CACCACCAGG CAAAGAGCGC AACAAACGGCA GGCAGCGCCA CCAGGCGAGT GTTTGATAACC AGGTTTCGCT ATGTTGAAC ACTACCGGGCG TTATGGCGGC ACCAAGCCCA CCGCCGCCCA TCAGAGCCGC GGACCACAGC CCCATCACCA GTGGCGTGCG CTGCTGAAAC CGCCGTTAA TCACCGAAGC ATCACCGCCT GAATGATGCC GATCCCCACC CCACCAAGCA GTGCGCTGCT AAGCAGCAGC GCACTTGCG GGTAAGCTC ACGCATCAAT	5940 6000 6060 6120 6180
25	GCACCGACGG CAATCAGCAA CAGACTGATG GCGACACTGC GACGTTCGCT GACATGCTGA TGAAGCCAGC TTCCGGCCAG CGCCAGCCCG CCCATGGTAA CCACCGGCAG AGCGGTCGAC CCGGACGGGA CGCTCCTGCG CCTGATACAG AACGAATTGC TTGCAGGCAT CTCATGAGTG TGTCTTCCCCT TTTTCCGCT GAGGTCACTG CGTGGATGGA GCGCTGGCGC CTGCTGCGCG ACGGCGAGCT GCTCACCACC CACTCGAGCT GGATACTTCC CGTCCGCCAG GGGGACATGC	6240 6300 6360 6420 6480
30	CGGCGATGCT GAAGGTCGCG CGCATTCCCG ATGAAGAGGC CGGTTACCGC CTGTTGACCT GGTGGGACGG GCAGGGCGCC GCCCGAGTCT TCGCCTCGGC GGCGGGCGCT CTGCTCATGG AGCGCGCGTC CGGGGCCGGG GACCTTGCAC AGATAGCGTG GTCCGCCAG GACGACGAGG CTTGCAGGAT CTATGATTCC CTTTGTCAAC AGCAATGGAT CACTGAAAAT GGTTCAATGA TCACATTAAG TGGTATTCAA TATTTTCATG AAATGGGAAT TGACGTTCCCT TCCAAACATT	6540 6600 6660 6720 6780
35	CACGTAAAAT CTGTTGTGCG TGTTTAGATT GGAGTGAACG CCGTTTCCAT TTAGGTGGGT ACGTTGGAGC CGCATTATTT TCGCTTATG AATCTAAAGG GTGGTTAACT CGACATCTTG GTTACCGTGA AGTTACCATC ACGGAAAAAG GTTATGCTGC TTTAAGACC CACTTCACA TTTAAGTTGT TTTCTAATC CGCATATGAT CAATTCAAGG CGGAATAAGA AGGCTGGCTC	6840 6900 6960 7020

5	TGCACCTTGG TGATCAAATA ATTGATAGC TTGTCGTAAT AATGGCGGCA TACTATCAGT AGTAGGTGTT TCCCTTTCTT CTTAGCGAC TTGATGCTCT TGATCTTCCA ATACGCAACC TAAAGTAAAA TGCCCCACAG CGCTGAGTGC ATATAATGCA TTCTCTAGTG AAAAACCTTG TTGGCATAAA AAGGCTAATT GATTTCGAG AGTTTCATAC TGTTTTCTG TAGGCCGTGT ACCTAAATGT ACTTTTGCTC CATCGCGATG ACTTAGTAAA GCACATCTAA AACCTTTAGC	7080 7140 7200 7260 7320
10	GTTATTACGT AAAAATCTT GCCAGCTTTC CCCTTCTAAA GGGCAAAAGT GAGTATGGTG CCTATCTAAC ATCTCAATGG CTAAGCGTC GAGCAAAGCC CGCTTATTTT TTACATGCCA ATACAATGTA GGCTGCTCTA CACCTAGCTT CTGGGCGAGT TTACGGGTTG TTAAACCTTC GATTCCGACC TCATTAAGCA GCTCTAATGC GCTGTTAAC CACTTACTTT TATCTAATCT AGACATCATT AATTCTAAT TTTTGTGAC ACTCTATCAT TGATAGAGTT ATTTTACAC	7380 7440 7500 7560 7620
15	TCCCTATCAG TGATAGAGAA AAGTGAATG AATAGTTCGA CAAAGATCGC ATTGGTAATT ACGTTACTCG ATGCCATGGG GATTGGCCTT ATCATGCCAG TCTTGCCAAC GTTATTACGT GAATTTATTG CTCGGAAGA TATCGCTAAC CACTTGGCG TATTGCTTGC ACTTTATGCG TTAATGCAGG TTATCTTGC TCCTTGGCTT GGAAAAATGT CTGACCGATT TGGTCGGCGC CCAGTGCTGT TGGTGTGATT AATAGGGCA TCGCTGGATT ACTTATTGCT GGCTTTTCA	7680 7740 7800 7860 7920
20	AGTGCCTTT GGATGCTGTA TTTAGGCCGT TTGCTTTCAG GGATCACAGG AGCTACTGGG GCTGTCGCGG CATCGGTCA TGCCGATACC ACCTCAGCTT CTCAACGCGT GAAGTGGTTC GGTTGGTTAG GGGCAAGTTT TGGGCTTGGT TTAATAGCGG GGCCTATTAT TGGTGGTTT GCAGGAGAGA TTTCACCGCA TAGTCCCTTT TTTATCGCTG CGTTGCTAAA TATTGTCACT TTCCTTGTGG TTATGTTTG GTTCCGTGAA ACCAAAAATA CACGTGATAA TACAGATACC	7980 8040 8100 8160 8220
25	GAAGTAGGGG TTGAGACGCA ATCGAATTG GTATACATCA CTTTATTTAA AACGATGCC ATTTTGTGA TTATTTATT TTCAGCGAA TTGATAGGCC AAATTCCGC AACGGTGTGG GTGCTATTAA CCGAAAATCG TTTGGATGG AATAGCATGA TGGTGGCTT TTCATTAGCG GGTCTTGGTC TTTTACACTC AGTATTCCAA GCCTTGTGG CAGGAAGAAT AGCCACTAAA TGGGGCGAAA AAACGGCAGT ACTGCTCGAA TTTATTGCAAG ATAGTAGTGC ATTTGCCTTT	8280 8340 8400 8460 8520
30	TTAGCGTTA TATCTGAAGG TTGGTTAGAT TTCCCTGTT TAATTTATT GGCTGGTGGT GGGATCGCTT TACCTGCATT ACAGGGAGTG ATGTCTATCC AAACAAAGAG TCATGAGCAA GGTGCTTAC AGGGATTATT GGTGAGCCTT ACCAATGCAA CCGGTGTTAT TGGCCCATT CTGTTACTG TTATTTATAA TCATTCACTA CCAATTGGG ATGGCTGGAT TTGGATTATT GGTTTAGCGT TTTACTGTAT TATTATCCTG CTATCGATGA CCTTCATGTT AACCCCTCAA	8580 8640 8700 8760 8820
35	GCTCAGGGGA GTAAACAGGA GACAAGTGCT TAGTTATTTC GTCACCAAAT GATGTTATTC CGCGAAATAT AATGACCCTC TTGATAACCC AAGAGGGCAT TTTTACGAT AAAGAAGATT TAGCTTCAA TAAAACCTAT CTATTTATT TATCTTCAA GCTCAATAAA AAGCCGCGGT AAATAGCAAT AAATTGGCCT TTTTATCGG CAAGCTCTT TAGGTTTTTC GCATGTATTG	8880 8940 9000 9060

5	CGATATGCAT AAACCAGCCA TTGAGTAAGT TTTTAAGCAC ATCACTATCA TAAGCTTAA GTTGGTTCTC TTGGATCAAT TTGCTGACAA TGGCGTTAC CTTACCAAGTA ATGTATTCAA GGCTAATTT TTCAAGTTCA TTCCAACCAA TGATAGGCAT CACTTCTTGG ATAGGGATAA GGTTTTATT ATTATCAATA ATATAATCAA GATAATGTTA AAATATACTT TCTAAGGCAG ACCAACCATT TGTTAAATCA GTTTTGTTG TGATGTAGGC ATCAATCATA ATTAATTGCT	9120 9180 9240 9300 9360
10	GCTTATAACA GGCACTGAGT AATTGTTTT TATTTTTAAA GTGATGATAA AAGGCACCTT TGGTCACCAA CGCTTTCCC GAGATCCTCT GCGACACCGC CGCTCGTCTG CACGCCCGC GGTCCGGACC GCCGCCCGAT CTCCATCCGC TACAGGAATG GTTCCAGCCG CTTTCCGGT TGGCCGCTGA GCACGCGGCA CTTGCGCCCG CCGCCAGCGT AGCGCGCCAA CTTCTGGCGG CGCCGCGCGA GGTGTGCCCG CTCCACGGCG ACCTGCACCA CGAGAACGTG CTCGACTTCG	9420 9480 9540 9600 9660
15	GCGACCGCGG CTGGCTGGCC ATCGACCCGC ACGGACTGCT CGCGAGCGC ACCTTCGACT ATGCCAACAT CTTCACGAAT CCCGATCTCA GCGACCCCGG TCGCCCGCTT GCGATCCTGC CGGGCAGGCT GGAGGCTCGA CTCAGCATTG TGGTCGCGAC GACCGGGTTT GAGCCCGAAC GGCTTCTTCG CTGGATCATT GCATGGACGG GCTTGTGGC AGCCTGGTTC ATCGGGGACG GCGACGGCGA GGGCGAGGGC GCTGCGATTG ATCTGGCCGT AAACGCCATG GCACGCCGGT	9720 9780 9840 9900 9960
20	TGCTTGACTA GCGCGGTAC CGATCTCAC TGGTCGTCGA GCTAGGTCAG GCCGTGTCGG GCGTGATCCG CTGGAAGTCG TTGCGGGCCA CACCCGCCGC CTCGAAGCCC TGCACCAGGC CGGCATCGTG GTGTGCGTGG CCGAGGGACT ATGGAAGGTG CCGGACGATC TGCCCGAGCA GGGCCGCCGC TATGACGCC AGCGTCTTGG TGGCGTGACG GTGGAGCTGA AATCGCACCT GCCCATCGAG CGGCAGGCC CGGTGATCGG TGCCACCTGG CTTGACCAGC AGTTGATCGA	10020 10080 10140 10200 10260
25	CGGTGGCTCG GGCTTGGCG ACCTGGGCTT TAGCAGTGAG GCCAAGTAGG CGATACAGCA GCGCGCGGAC TTCCTGGCG AACAGGGACT GGCGAGCGG CGCGGGCAGC GCGTGATCCT CACCGGAATC TGCTGGCAG CAGCGGGCTC GGGAACTGGC GCAGGCCCG AAGGACATTG CCGCCGATAC CGGCCTGGAG CATCGCCCCG TGGCGACGG CCAGCGCGTT GCCGGCGTCT ACCGGGGCC CGTCATGCTC GCCAGCGGGC GAAATGGGAT GCTTGATGAC GCCAAGGGGT	10320 10380 10440 10500 10560
30	CCAGCCTCGT GCGGTGGAAG CCCATCGAAC AGCGGTTGG GGAGCAGCTC GCCGCGACGG TGCGCGGTGG CGGCGTGTCT TGGGAGATTG GACGACAGCG TGGGCCGGCC CCTGTCTCTT GATCAGATCT TGATCCCCCTG CGCCATCAGA TCCTTGGCGG CAAGAAAGCC ATCCAGTTA CTTTCAGGG CTTCCAACC TTCCAGAGG GCGCCCGAGC TGGCAATTCC GGTCGCTTG CTGTCCATAA AACCGCCAG TCTAGCTATC GCCATGTAAG CCCACTGCAA GCTACCTGCT	10620 10680 10740 10800 10860
35	TTCTCTTGC GCTTGCGTTT TCCCTTGTCC AGATAGCCCA GTAGCTGACA TTCATCCGGG GTCAGCACCG TTTCTGCGGA CTGGCTTCT ACGTGTTCCG CTTCTTTAG CAGCCCTTGC GCCCTGAGTG CTTGCGGCAG CGTGAAGCTT TCTCTGAGCT GTAACAGCCT GACCGCAACA AACGAGAGGA TCGAGACCAC CCGCTCCAGA TTATCCGGCT CCTCCATGCG TTGCCTCTCG	10920 10980 11040 11100

5	GCTCCTGCTC CGGTTTCCA TGCCTATGG AACTCCTCGA TCCGCCAGCG ATGGGTATAA	11160
	ATGTCGATGA CGCGCAAGGC TTGGGCTAGC GACTCGACCG GTTCGCCGGT CAGCAACAAC	11220
	CATTTCACG GGGTCTCACCC CTTGGCGGG TTAATCTCCT CGGCCAGCAC CGCGTTGAGC	11280
	GTGATATTCC CCTGTTTAG CGTGATGCGC CCACTGCGCA GGCTCAAGCT CGCCTGCAG	11340
	GCTGGTCGAT TTTTACGTTT ACCCGCGTTA TCCACCACGC CCTTTGCAG AATGCTGATC	11400
10	TGATAGCCAC CCAACTCCGG TTGGTTCTTC AGATGGTCGA TCAGATAACAA CCCAGACTCT	11460
	ACGTCCCTGCGT GGAGCGCACC ACGAAGCGCT CGTTATGCGC CAGCCTGTCC	11520
	TGCAGATAAG CATGAATATC GGCTTCGCGG TCACAGACCG CAATCACGTT GCTCATCATG	11580
	CTGCCCATGC GTAACCGGGCT AGTTGCGGCC GCTGCCAGCC ATTTGCCACT CTCCTTTCA	11640
	TCCGCATCGG CAGGGTCATC CGGGCGCATC CACCACTCCT GATGCAGTAA TCCTACGGTG	11700
15	CGGAATGTGG TGGCCTCGAG CAAGAGAACG GAGTGAACCC ACCATCCGCG GGATTATCC	11760
	TGAATAGAGC CCAGCTTGCC AAGCTTTGCG GCGACCTGGT GGCGATAACT CAAAGAGGTG	11820
	GTGTCCCTCAA TG GCCAGCAG TTCGGAAAC TCCTGAGCCA ACTTGACTGT TTGCATGGCG	11880
	CCAGCCTTTC TGATCGCCTC GGCAGAAACG TTGGGATTGC GGATAAAATCG GTAAGCGCCT	11940
	TCCTGCATGG CTTCACTACC CTCTGATGAG ATGGTTATTG ATTTACCAGA ATATTTGCC	12000
20	AATTGGCGG CGACGTTAAC CAAGCGGGCA GTACGGCGAG GATCACCCAG CGCCGCCGAA	12060
	GAGAACACAG ATTTAGCCC GTCGGCCGCA CGATGAAGAG CAGAAGTTAT CATGAACGTT	12120
	ACCATGTTAG GAGGTCACAT GGAAGATCAG ATCCTGGAAA ACGGGAAAGG TTCCGTTCGA	12180
	ATTGCATGCG GATCCGGGAT CAAGATCTGA TCAAGAGACA GGTACCAATT GTTGAAGACG	12240
	AAAGGGCCTC GTGATACGCC TATTTTATA GGTAAATGTC ATGATAATAA TGGTTTCTTA	12300
25	GACGTCAGGT GGCACCTTTTC GGGGAAATGT GCGCGGAACC CCTATTTGTT TATTTTCTA	12360
	AATACATTCA AATATGTATC CGCTCATGAG ACAATAACCC TGATAAAATGC TTCAATAATA	12420
	TTGAAAAAGG AAGAGTATGA GTATTCAACA TTTCCGTGTC GCCCTTATTG CCTTTTTGC	12480
	GGCATTTCGC CTTCTGTGTT TTGCTCACCC AGAAACGCTG GTGAAAGTAA AAGATGCTGA	12540
	AGATCAGTTG GGTGCACGAG TGGTTACAT CGAACTGGAT CTCAACAGCG GTAAGATCCT	12600
30	TGAGAGTTT CGCCCCGAAG AACGTTTCC AATGATGAGC ACTTTAAAG TTCTGCTATG	12660
	TGGCGCGGTA TTATCCCCTG TTGACGCCGG GCAAGAGCAA CTCGGTCGCC GCATACACTA	12720
	TTCTCAGAAT GACTTGGTTG AGTACTTGGC AAAACTGATCT AAATGTTAG CCCAGTCATC	12780
	ATACTTCACC GATGCCAACG CATTAAAAT AGCATCACGA TCGGCTTGC TGAATTCTT	12840
	ATTTAAAACA TCCTTGTATT TTTCAAAAGC AGCGAGAGCT TCATTCACAT TGCCGATT	12900
35	CTTACCTTTA GACTTATCAG CAAGTTCTG TGCCATTTC GAATATTTT CACCATATT	12960
	TTCAGTCAGC GTTTGATAAA AGCTAACTGT TGCATCAACA GCATCCTTAA TCTGTGAATT	13020
	AAGGAGAGTTA TTCTGTGCTT TTTCAAAATT TTCTTCAGCT TCATGAACAC GAGCGATACC	13080
	GGCATTACGA TTATTACTGA CCTGAGAAAT AGCCTTCTGG ATCTGAGTTA TATCAGCATT	13140

5	TATCCGGTTA ATACGTGTTT CTGATGCTGT TACCTGTTT TGTTTTCTT CTCTAACTT 13200
	ACCGGCCCA ACCCGTCGTC TGTTGCTTC AAAAAAAGGA CGGTTCTGAA GCGGATCATT 13260
	GGCTCTTGGT GATAGTTTT TGACCAAGCTC ATCCAGTTCT TTATATTTAG CGGATGCCTG 13320
	AGCCAGTTCA TTTCGTTTC CAGCGAGCGT TTTCATTCT GCATCACGGG CATGGATACT 13380
	GGAGCTTAAA CGAGAATTGA GAGTCTTAAT CTCTCCATCC ATTTTCACCA CTTCAGATTG 13440
10	TGCAGCAGAA AGTTTTTTT GGGCGATCTC AACAGCTTTA GCTTCTTCAC TCAATGCAGC 13500
	CAGTCGTTTC TCTTCAGCTT CAGCCAGTTT CAACTGGCGT TCTGTTTCAG CCTTCTCCCG 13560
	TTCAATCTCT TTACGTGTT GTTCTGCTTC CTGAAAAGCC TTTTCTGCTG CTTCCGCTTC 13620
	TTTACGGGCT TTTTCTTCTG CTTTCGCAAG GCGCAAACGC TCTGCTTCCG CCTGCATAGC 13680
	TGCATTATTA GCATGAGCAA GCTCTGTTGC TGAAGGCGTA CGTGAGGCAT TGTGACGAAG 13740
15	AGCCTCATTC ACGATATCCT TCAGGCGCTG AGTCAGCGCA TCCCTGTTTGC CCTTGCTTT 13800
	CGCCTGTGCT TCCGCTGCAG CTTTGCCCCG GGCAGCCTGC TCTGCCTGTG TTTTCTTAA 13860
	TTGAGCAGTA GACCATTAG CAGTTGCATG AATAGCTGCA GAACTTCAC TTTTACTGCC 13920
	TCCTTTCCA CCTCCGCCGC CAGAGCCACT CCCGTCAGGA GTACCATTCA AAAGAGTAAT 13980
	AATTACCTGT CCCTTATCAT CATAAGGAAC ACCATCTTA TAGTACGCTA CCGCGGTTTC 14040
20	CATTATAAAA TCCTCTTGA CTTTAAAC AATAAGTTAA AAATAAAATAC TGTACATATA 14100
	ACCACTGGTT TTATATACAG CATAAAAGCT ACGCCGCTGC ATTTCCTCG TCAAGACTGT 14160
	GGACTTCCAT TTTTGTGAAA ACGATCAAAA AAACAGTCTT TCACACCACG CGCTATTCTC 14220
	GCCCCATGCC ACAAAAACCA GCACAAACAT TACCGTTCTC AGACCTCATT ATGTTTACT 14280
	GAAACTATGA GATGAGACAT CTATGGGACA CTGTCACCTT ATGGCATGGC ACACACTCCG 14340
25	GGACGCACTA AAAATGACAG GCAGATCGCG TTCACAGTTT TACCGTGATA TGCGCGGAGG 14400
	CCTTGTCACT TACCGTACCG GCAGGGACGG ACGACGGGAG TTTGAAACCA GTGAACGTGAT 14460
	CCGGGCATAC GGCGAATTAA AGCAGAACAT GACACCAGAA AGGCACAGTG AGGGACATGC 14520
	AGAAAATCCA CATGATCAGC AGACAGAACG CATTCTCCGG GAACTGAATG AGCTGAAACA 14580
	ATGCCTGACG CTGATGCTTG AGGATAAACCA GGCACAGGAT ATGGATCGCA GACGCCAGGA 14640
30	AGCAGAACGG GAACAGCTAC AAAATGAGAT AGCCCAGCTC AGGCAGGCAC TGGAACTGG 14700
	AAAGAAACGG GGATTCTGGT CCAGGTTGTT CGGTCGCTGA ACGCTGTCAG AGACTGATGA 14760
	TAAAATAGTC TTCGGATAAT AACTCACCAGA GAATAAAATAC TTTAAGGTAG GGAGACACTC 14820
	ATGAGACGTA CCGGAAACAA ACTTTGTCTT ATGCCATGA TAACAGCAAC AGTAGCTCTC 14880
	ACAGCCTGTA CCCCAAAGGG CAGCGTGGAA CAACATACCC GGCATTACGT ATATGCTTCT 14940
35	GATGACGGTT TTGATCCCAA CTTTCCACC CAAAAAGCCG ACACAACACG AATGATGGTG 15000
	CCTTTTTTC GGCAGTTCTG GGATATGGGA GCTAAAGACA AAGCGACAGG AAAATCACGG 15060
	AGTGATGTGC AACAAACGCAT TCAGCAGTTT CACAGCCAAG AATTTTAAA CTCACTCCGG 15120
	GGCACAAACTC AATTGCGGG TACTGATTAC CGCAGCAAAG ACCTTACCCC GAAAAAAATCC 15180

5	AGGCTGCTGG CTGACACGAT TTCTGCGGT TATCTCGATG GCTACGAGGG CAGACAGTAA	15240
	GTGGATTAC CATAATCCCT TAATTGTACG CACCGCTAAA ACGCGTTCAG CGCGATCACG	15300
	GCAGCAGACA GGTAAAAATG GCAACAAACC ACCCTAAAAA CTGCGGATC GCGCCTGATA	15360
	AATTTAACC GTATGAATAC CTATGCAACC AGAGGGTACA GGCCACATTA CCCCCACTTA	15420
	ATCCACTGAA GCTGCCATT TTCATGGTTT CACCATCCC GCGAAGGGCC ATCCAGCGTG	15480
10	CGTTCCGTGA TTTCCGGCTG ACGCTCCCGT TCTAGGGATA ACACATGTTG GCGCTCCTGT	15540
	ATCAGCCGTT CCTCTCTTAT CTCCAGTTCT CGCTGTATAA CTGGCTCAAG CGTTCTGTCT	15600
	GCTCGCTCAA GTGTTGCACC TGCTGACTCA ACTGCATGAC CCGCTCGTTC AGCATCGCGT	15660
	TGTCCCGTTG CGTAAGCGAA AACATCTTCT GCAATTCCAC GAAGGCCTC TCCCATTCGC	15720
	TCAGCCGCTG CATATAGTCC TGTTGCAGCT GCTCTAAGGC GTTCAGCAAA TGTGTTCCA	15780
15	GCTCTGTCAC TCTGTGTCAC TCCTTCAGAT GTACCCACTC TTTCCCTGA AAGGAAATCA	15840
	CCTCCGCTGA TTTCCCGTAC GGAAGGACAA GGAATTTCCT GTTCCCGTCC TGCACAAACT	15900
	CCACGCCCCA TGTCTTCGCG TTCAGTTCT GCAATGTCTC TTCTGCTTC CTGATTTCTT	15960
	CCAGGTTCGC CTGTATCCTC CCTCCAAGAT ACCAGAGCGT CCCGCCACTC GCGGTAAACA	16020
	GGAGAAAGAC TATCCCCAGT AACATCATGC CCGTATTCCC TGCCAGCTTT AACACGTCCC	16080
20	TCCTGTGCTG CATCATCGCC TCTTCACCC CTTCCCGGTG TTTTCCAGC GATTCCCTTG	16140
	TCGAGGCTGT GAACAGGGCT ATAGCGTCTC TGATTTCTG CTCGTTGAT GTCACAGCCT	16200
	CGCTTACAGA TTGCGCGAGC CTCCTGAACG CGTTGTTCAAG CATTTCCTCT GTAGATTCGG	16260
	CTCTCTCTTT CAGCTTTTC TCGAACTCCG CGCCCGTCTG CAAAAGATTG CTCATAAAAT	16320
	GCTCCTTCA GCCTGATATT CTTCCCGCCG TTCGGATCTG CAATGCTGAT ACTGCTTCGC	16380
25	GTCACCCCTGA CCACCTCCAG CCCCCCCTCA GTGAGCGCCT GAATCACATC CTGACGGCCT	16440
	TTTATCTCTC CGGCATGGTA AAGTGCATCT ATACCTCGCG TGACGCCCTC AGCAAGCGCC	16500
	TGTTTCGTTT CAGGCAGGTT ATCAGGGAGT GTCAGCGTCC TGCGGTTCTC CGGGCGTTC	16560
	GGGTCATGCA GCCCGTAATG GTGATTTAAC AGCGTCTGCC AAGCATCAAT TCTAGGCCTG	16620
	TCTGCGCGGT CGTAGTACGG CTGGAGGCGT TTTCCGGTCT GTAGCTCCAT GTTCGGAATG	16680
30	ACAAAATTCA GCTCAAGCCG TCCCTGTCC TGGTGCTCCA CCCACAGGAT GCTGTACTGA	16740
	TTTTTTTCGA GACCGGGCAT CAGTACACGC TCAAAGCTCG CCATCACTTT TTCACGTCCT	16800
	CCCGGGCGCA GCTCCTCTC CGCGAACGAC AGAACACCGG ACGTGTATTT CTTCGCAAAT	16860
	GGCGTGGCAT CGATGAGTTC CCGGACTTCT TCCGGTATAC CCTGAAGCAC CGTTGCGCCT	16920
	TCGCGGTTAC GCTCCCTCCC CAGCAGGTAA TCAACCGGAC CACTGCCACC ACCTTTCCC	16980
35	CTGGCATGAA ATTTAACTAT CATCCCGCGC CCCCTGTTCC CTGACAGCCA GACGCAGCCG	17040
	GCGCAGCTCA TCCCCGATGG CCATCAGTGC GGCCACCACC TGAACCCGGT CACCGGAAGA	17100
	CCACTGCCCG CTGTTCACCT TACGGGCTGT CTGATTCAAGG TTATTTCCGA TGGCGGCCAG	17160
	CTGACGCAGT AACGGCGGTG CCAGTGTCTGG CAGTTTCCG GAACGGGCAA CCGGCTCCCC	17220

5	CAGGCAGACC CGCCGCATCC ATACCGCCAG TTGTTTACCC TCACAGCGTT CAAGTAACCG GGCATGTTCA TCATCAGTAA CCCGTATTGT GAGCATTCTC TCGCGTTCA TCGGTATCAT TACCCCATGA ACAGAAATCC CCCTTACACG GAGGCATCAG TGACTAAACA GGAAAAAACC GCCCTTAACA TGGCCCGCTT TATCAGAAGC CAGACATTAA CGCTGCTGGA GAAGCTCAAC GAACTGGACG CAGATGAACA GGCGATATT TGTGAATCGC TTCACGACCA CGCGATGAG	17280 17340 17400 17460 17520
10	CTTTACCGCA GCTGCCTCGC ACGTTTCGGG GATGACGGTG AAAACCTCTG ACACATGCAG CTCCCGGAGA CGGTCACAGC TTGCTGTGA GCGGATGCCG GGAGCTGACA AGCCCGTCAG GGCGCGTCAG CAGGTTTAG CGGGTGTGCG GGCGCAGCCC TGACCCAGTC ACGTAGCGAT AGCGGAGTGT ATACTGGCTT ACCATGCGG CATCAGTGCG GATTGTATGA AAAGTACGCC ATGCCGGGTG TGAAATGCCG CACAGATGCG TAAGGAGAAA ATGCACGTCC AGGCGCTTT	17580 17640 17700 17760 17820
15	CCGCTTCCTC GCTCACTGAC TCGCTACGCT CGGTCGTTCG ACTGCGGCGA GCGGTACTGA CTCACACAAA AACGGTAACA CAGTTATCCA CAGAACATCAGG GGATAAGGCC GGAAAGAAC TGTGAGCAAA AGACCAGGAA CAGGAAGAAC GCCACGTAGC AGGCGTTTT CCATAGGCTC CGCCCCCTG ACGAGCATCA CAAAAATAGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA GGACTATAAA GCTACCAGGC GTTCCCCCT GGAAGCTCCC TCGTGCCTC TCCTGTTCCG	17880 17940 18000 18060 18120
20	ACCCCTGCCGC TTACCGGATA CCTGTCCGCC TTTCTCCCTT CGGGAAAGCGT GGCGCTTTCT CATAGCTCAC GCTGTTGGTA TCTCAGTTCG GTGTAGGTCG TTCGCTCAA GCTGGGCTGT GTGCACGAAC CCCCCGTTCA GCCCGACCGC TGCGCCTTAT CCGGTAACTA TCGTCTTGAG TCCAACCCGG TAAGGCACGC CTTAACGCCA CTGGCAGCAG CCACTGGTAA CCGGATTAGC AGAGCGATGA TGGCACAAAC GGTGCTACAG AGTTCTTGAA GTAGTGGCCC GACTACGGCT	18180 18240 18300 18360 18420
25	ACACTAGAAG GACAGTATTT GGTATCTGCG CTCTGCTGAA GCCAGTTACC TTCGGAAAAA GAGTTGGTAG CTCTTGATCC GGCAAACAAA CCACCGTTGG TAGCGGTGGT TTTTTGTTT GCAAGCAGCA GATTACGGCG AGAAAAAAAG GATCTCAAGA AGATCCTTTA ATCTTTCTA CTGAACCGCG ATCCCCGTCA GTTTAGAAGA GGAGGATGGT GCGATGGTCC CTCCCTGAAC ATCAGGTATA TAGTTAGCCT GACATCCAAC AAGGAGGTTT ATCGCGAATA TTCCCACAAA	18480 18540 18600 18660 18720
30	AAATCTTTTC CTCATAACTC GATCCTTATA AAATGAAAAG AATATATGGC GAGGTTAAT TTATGAGCTT AAGATACTAC ATAAAAAAATA TTTTATTGCG CCTGTACTGC ACACCTATAT ATATATACCT TATAACAAAA AACAGCGAAG GGTATTATTT CCTTGTGTCA GATAAGATGC TATATGCAAT AGTGATAAGC ACTATTCTAT GTCCATATTG AAAATATGCT ATTGAATACA TAGCTTTAA CTTCATAAAG AAAGATTTT TCGAAAGAAG AAAAAACCTA AATAACGCC	18780 18840 18900 18960 19020
35	CCGTAGCAAA ATTAAACCTA TTTATGCTAT ATAATCTACT TTGTTGGTC CTAGCAATCC CATTGGATT GCTAGGACTT TTTATATCAA TAAAGAATAA TTAAATCCCT AACACCTCAT TTATAGTATT AAGTTTATTC TTATCAATAT AGGAGCATAG AA	19080 19140 19182

1. A system for transposing a transposable DNA sequence *in vitro*, the system comprising:

a Tn5 transposase modified relative to a wild type Tn5 transposase, the modified transposase comprising a change relative to the wild type Tn5 transposase that causes the modified transposase to bind to Tn5 outside end repeat sequences with greater avidity than the wild type Tn5 transposase, and a change relative to the wild type Tn5 transposase that causes the modified transposase to be less likely than the wild type transposase to assume an inactive multimeric form;

a donor DNA molecule comprising the transposable DNA sequence, the DNA sequence being flanked at its 5'- and 3'-ends by the Tn5 outside end repeat sequences; and

a target DNA molecule into which the transposable element can transpose.

2. A system as claimed in Claim 1 wherein the change that causes the modified transposase to bind with greater avidity is characterized as a substitution mutation at position 54 of the wild type transposase.

3. A system as claimed in Claim 2 wherein position 54 is a lysine.

4. A system as claimed in Claim 1 wherein the change that causes the modified transposase to be less likely to assume an inactive multimeric form is characterized as a substitution mutation at position 372 of the wild type transposase.

5. A system as claimed in Claim 4 wherein position 372 is a proline.

6. A system as claimed in Claim 1 wherein the modified transposase further comprises a substitution mutation at position 56 of the wild type transposase.

7. A system as claimed in Claim 6 wherein position 56 is an alanine.

8. A system as claimed in Claim 1 wherein the donor DNA molecule is flanked at its 5'- and 3'-ends by an 18 or 19 base pair flanking DNA sequence comprising nucleotide A at position 10, nucleotide T at position 11, and nucleotide A at position 12.

9. The system as claimed in Claim 8 wherein the flanking sequence further comprises a nucleotide at position 4 selected from the group consisting of A or T.

10. The system as claimed in Claim 8 wherein the flanking sequence further comprises a nucleotide at position 15 selected from the group consisting of G or C.

11. The system as claimed in Claim 8 wherein the flanking sequence further comprises a nucleotide at position 17 selected from the group consisting of A or T.

12. The system as claimed in Claim 8 wherein the flanking sequence further comprises a nucleotide at position 18 selected from the group consisting of G or C.

13. The system as claimed in Claim 8 wherein the flanking sequence has the sequence 5'-CTGTCTCTTATACACATCT-3'.

14. The system as claimed in Claim 8 wherein the flanking sequence has the sequence 5'-CTGTCTCTTATACAGATCT-3'.

15. A Tn5 transposase modified relative to a wild type Tn5 transposase, the modified transposase comprising:

a change relative to the wild type Tn5 transposase that causes the modified transposase to bind to Tn5 outside end repeat sequences of a donor DNA with greater avidity than the wild type Tn5 transposase; and

a change relative to the wild type Tn5 transposase that causes the modified transposase to be less likely than the wild type transposase to assume an inactive multimeric form.

16. A modified Tn5 transposase as claimed in Claim 15 wherein the change that causes the modified transposase to bind with greater avidity is characterized as a substitution mutation at position 54 of the wild type transposase.

17. A modified Tn5 transposase as claimed in Claim 16 wherein position 54 is a lysine.

18. A modified Tn5 transposase as claimed in Claim 15 wherein the change that causes the modified transposase to be less likely to assume an inactive multimeric form is characterized as a substitution mutation at position 372 of the wild type transposase.

19. A modified Tn5 transposase as claimed in Claim 18 wherein position 372 is a proline.

20. A modified Tn5 transposase as claimed in Claim 15 further comprising a substitution mutation at position 56 of the wild type transposase.

21. A modified Tn5 transposase as claimed in Claim 20 wherein position 56 is alanine.

22. A genetic construct comprising a nucleotide sequence that can encode a Tn5 transposase that both has greater avidity for Tn5 outside end repeats and is less likely to assume an inactive multimeric form than a wild type Tn5 transposase.

23. A genetic construct as claimed in Claim 22 comprising a nucleotide sequence that encodes a lysine residue at amino acid 54 of the transposase.

24. A genetic construct as claimed in Claim 22 comprising a nucleotide sequence that encodes a proline residue at amino acid 372 of the transposase.

25. A genetic construct as claimed in Claim 22 comprising a nucleotide sequence that encodes a lysine residue at amino acid 54 of the transposase and a proline residue at amino acid 372 of the transposase.

26. A genetic construct as claimed in Claim 22 comprising the nucleotide sequence of SEQ ID NO:1.

27. A genetic construct comprising:  
a transposable DNA sequence flanked at its 5' and 3' ends by an 18 or 19 base pair flanking DNA sequence comprising nucleotide A at position 10, nucleotide T at position 11, and nucleotide A at position 12.

28. The construct of Claim 27 further comprising, at position 4 of the flanking sequence, a nucleotide selected from the group consisting of T or A.

29. The construct of Claim 27 further comprising, at position 15 of the flanking sequence, a nucleotide selected from the group consisting of G or C.

30. The construct of Claim 27 further comprising, at position 17 of the flanking sequence, a nucleotide selected from the group consisting of T or A.

31. The construct of Claim 27 further comprising, at position 18 of the flanking sequence, a nucleotide selected from the group consisting of G or C.

32. The construct as claimed in Claim 27 wherein the flanking sequence has the sequence 5'-CTGTCTCTTATACACATCT-3'.

33. The construct as claimed in Claim 27 wherein the flanking sequence has the sequence 5'-CTGTCTCTTATACAGATCT-3'.

34. A method for *in vitro* transposition, the method comprising the steps of:

combining a donor DNA molecule that comprises a transposable DNA sequence of interest, the DNA sequence of interest being flanked at its 5'- and 3'-ends by Tn5 outside end repeat sequences, with a target DNA molecule and a Tn5 transposase modified relative to wild type Tn5 transposase in a suitable reaction buffer at a temperature below a physiological temperature until the modified transposase binds to the outside end repeat sequences; and

raising the temperature to a physiological temperature for a period of time sufficient for the enzyme to catalyze *in vitro* transposition,

wherein the modified transposase comprises a change relative to the wild type Tn5 transposase that causes the modified transposase to bind to the Tn5 outside end repeat sequences with greater avidity than the wild type Tn5 transposase, and a change relative to the wild type Tn5 transposase that causes the modified transposase to be less likely than the wild type transposase to assume an inactive multimeric form.

35. A method as claimed in Claim 34 wherein the change that causes the modified transposase to bind with greater avidity is characterized as a substitution mutation at position 54 of the wild type transposase.

36. A method as claimed in Claim 35 wherein position 54 is a lysine.

37. A method as claimed in Claim 34 wherein the change that causes the modified transposase to be less likely to assume an inactive multimeric form is characterized as a substitution mutation at position 372 of the wild type transposase.

38. A method as claimed in Claim 37 wherein position 372 is a proline.

39. A method as claimed in Claim 34 wherein the modified transposase further comprises a substitution mutation at position 56 of the wild type transposase.

40. A method as claimed in Claim 39 wherein position 56 is an alanine.

41. A method as claimed in Claim 34 wherein the DNA sequence of interest is flanked at its 5'- and 3'-ends by an 18 or 19 base pair flanking DNA sequence comprising nucleotide A at position 10, nucleotide T at position 11, and nucleotide A at position 12.

42. The method as claimed in Claim 41 wherein the flanking sequence further comprises a nucleotide at position 4 selected from the group consisting of A or T.

43. The method as claimed in Claim 41 wherein the flanking sequence further comprises a nucleotide at position 15 selected from the group consisting of G or C.

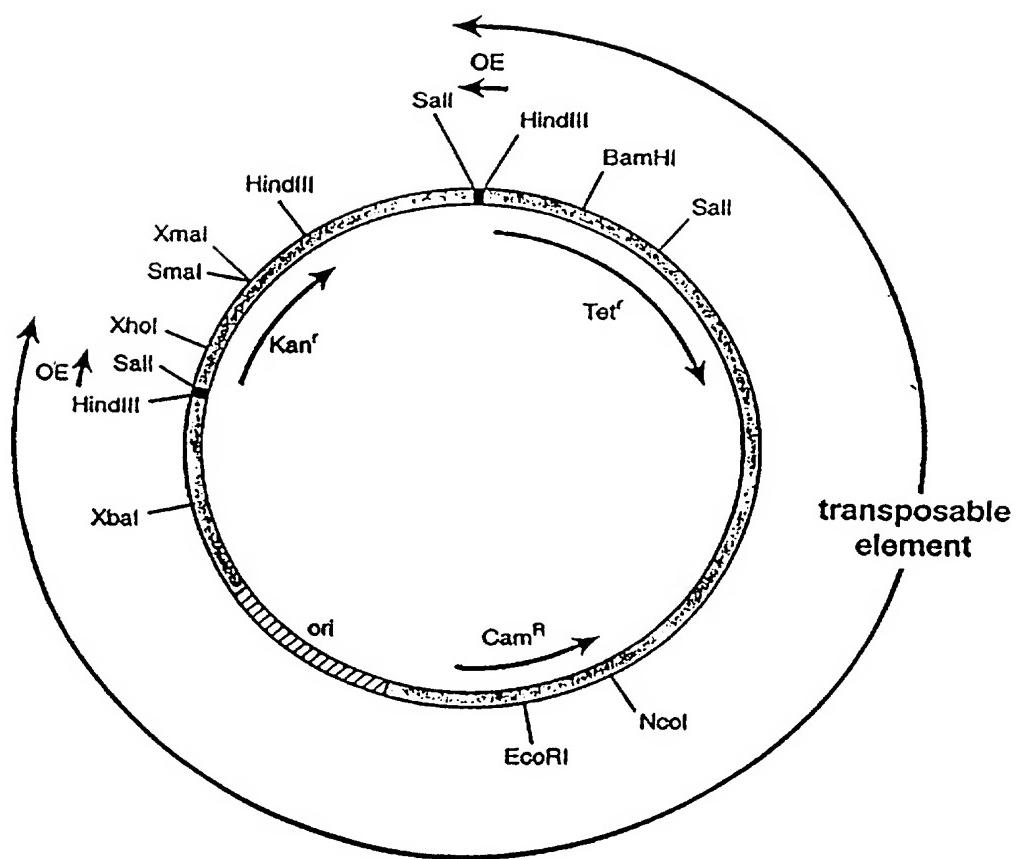
44. The method as claimed in Claim 41 wherein the flanking sequence further comprises a nucleotide at position 17 selected from the group consisting of A or T.

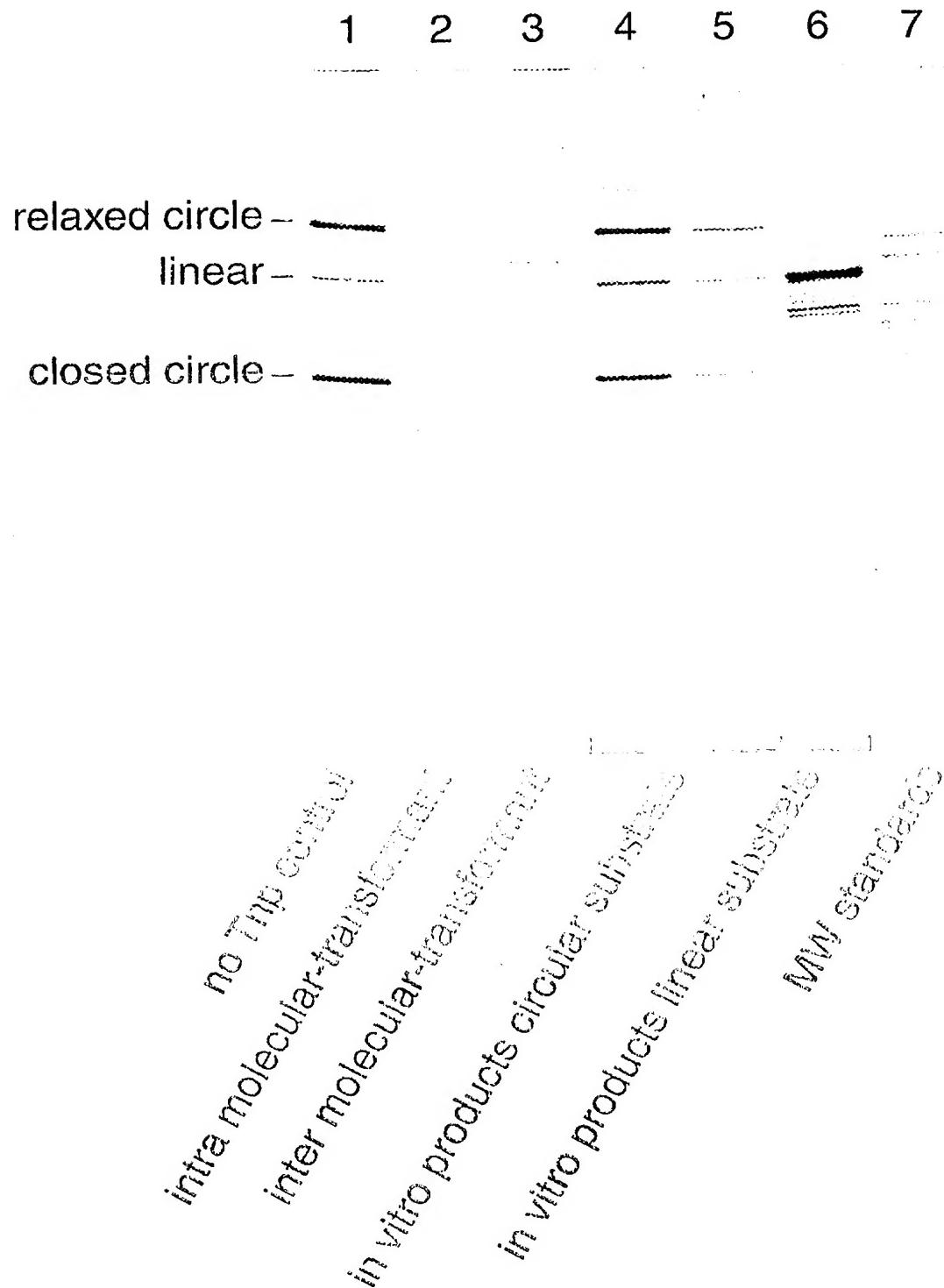
45. The method as claimed in Claim 41 wherein the flanking sequence further comprises a nucleotide at position 18 selected from the group consisting of G or C.

46. The method as claimed in Claim 41 wherein the flanking sequence has the sequence 5'-CTGTCTCTTATACACATCT-3'.

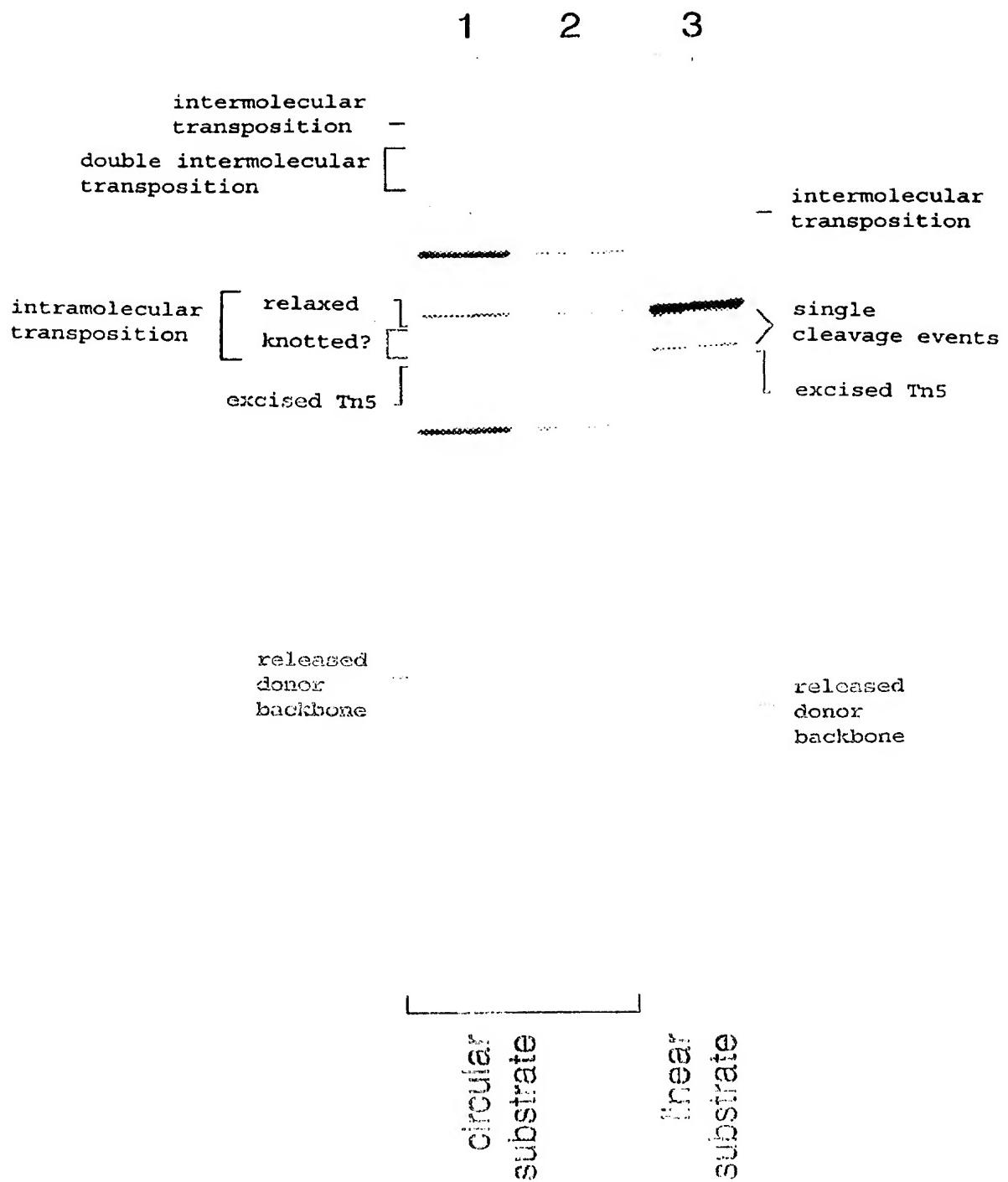
47. The method as claimed in Claim 41 wherein the flanking sequence has the sequence 5'-CTGTCTCTTATACAGATCT-3'.

1/8

**FIG 1**

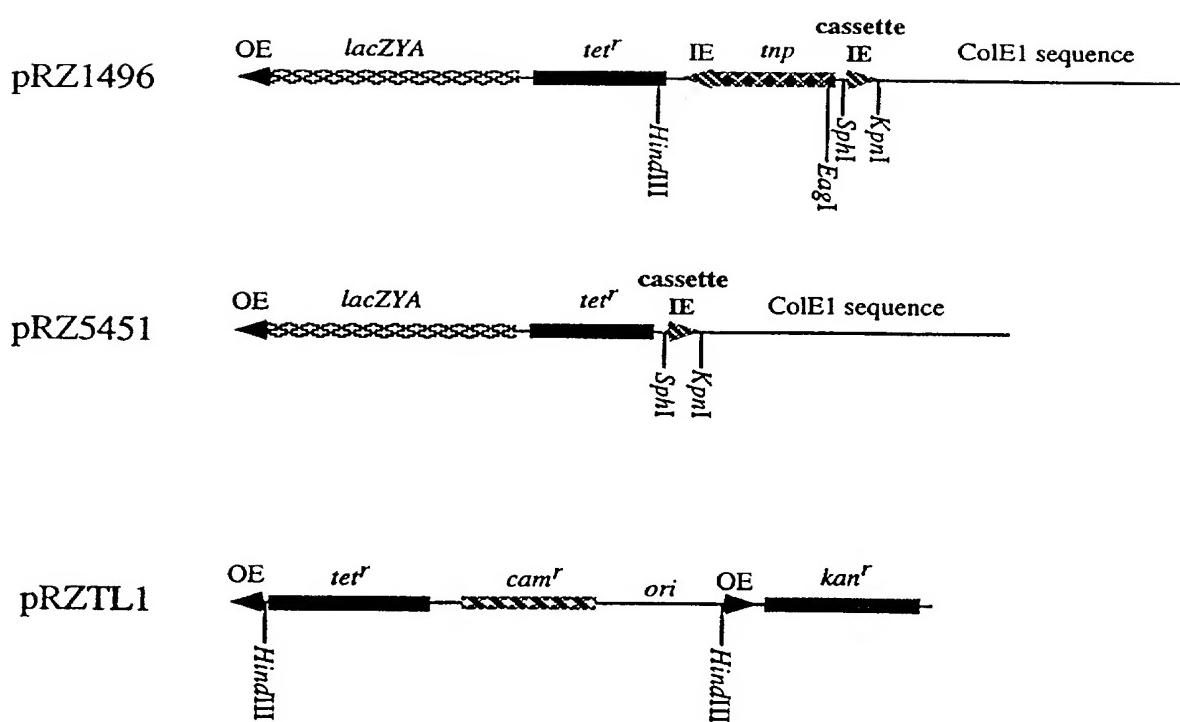


**FIG 2**  
**SUBSTITUTE SHEET (RULE 26)**

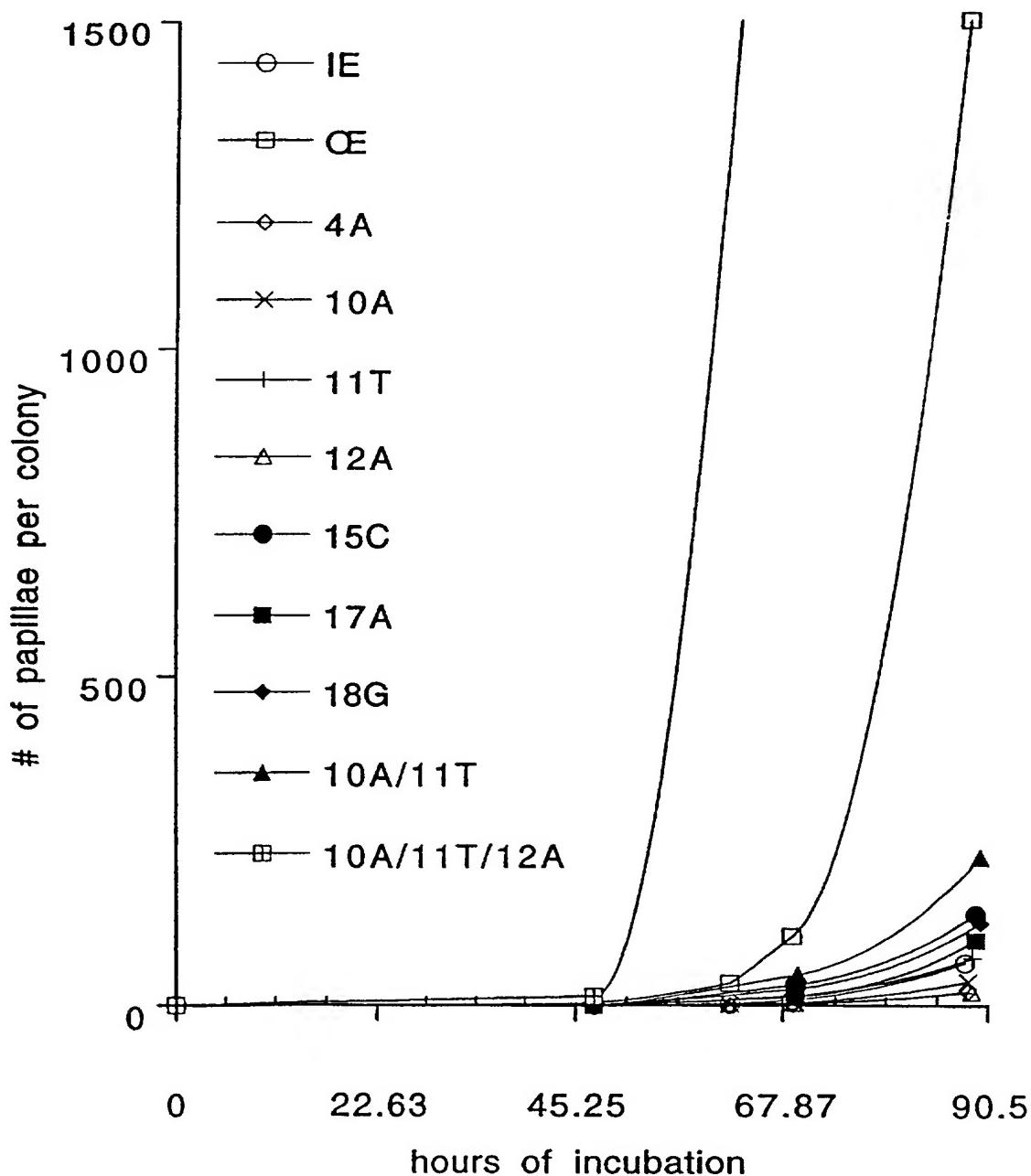


**FIG 3**  
**SUBSTITUTE SHEET (RULE 26)**

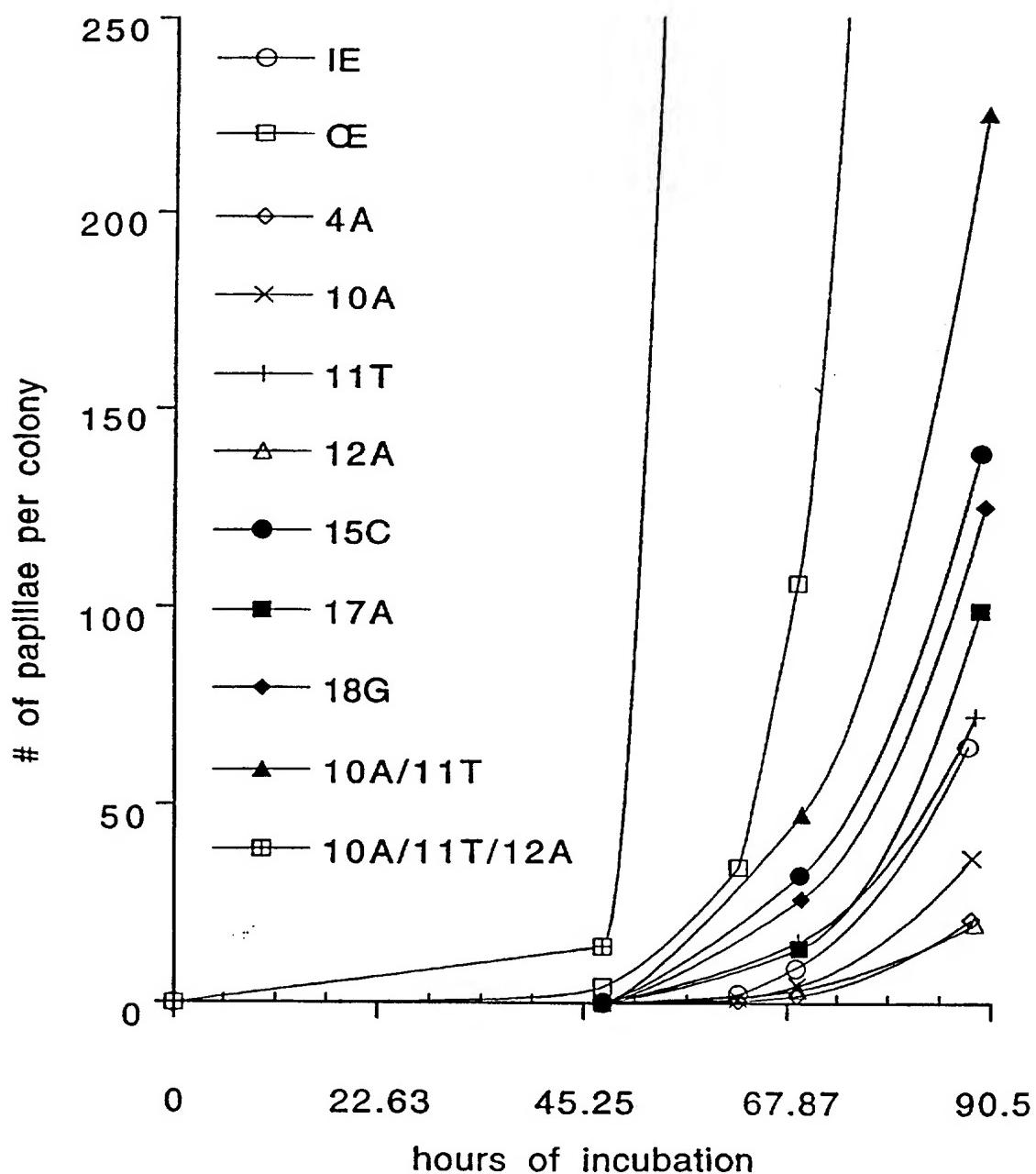
4 / 8

**FIG 4**

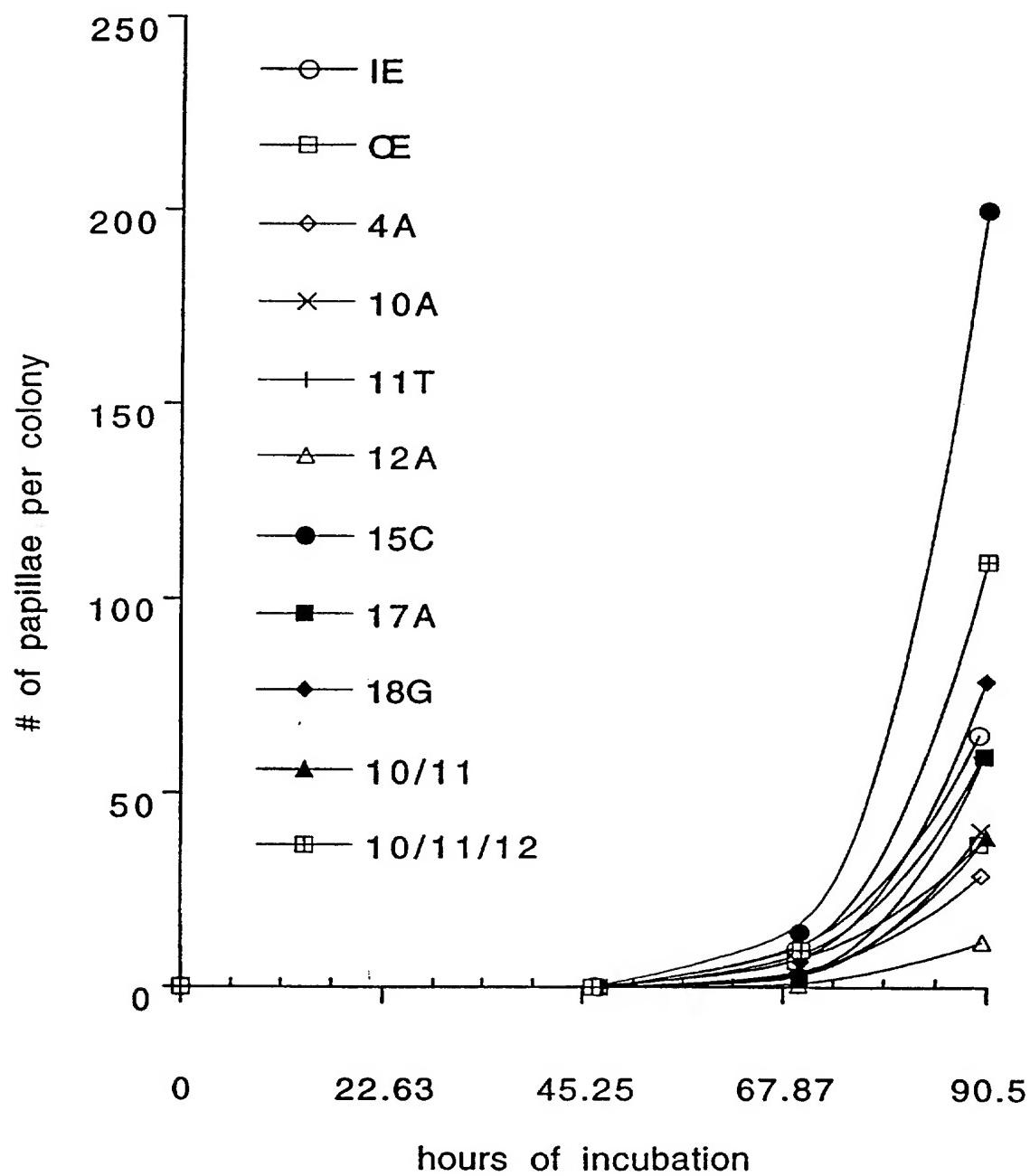
5/8

**a****Papillation of IE Mutants with EK54 Tnp****FIG 5**

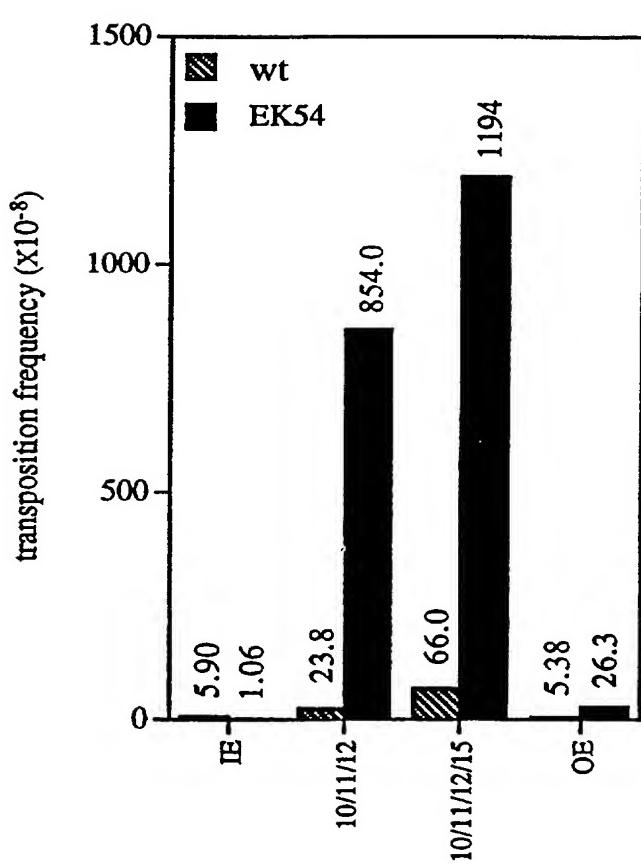
6/8

**b****Papillation of IE Mutants with EK54 Tnp****FIG 6**

7/8

**Papillation of IE Mutants with wt Tnp****FIG 7**

8/8

**a****FIG 8**

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 97/15941

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC 6 C12N15/55 C12N9/22 C12N15/90 C12N15/85

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>ZHOU M ET AL: "Three types of novel mutations in the NH-2-terminus of Tn5 transposase: Structure-function of transposase."</p> <p>KEYSTONE SYMPOSIUM ON TRANSPOSITION AND SITE-SPECIFIC RECOMBINATION: MECHANISM AND BIOLOGY, PARK CITY, UTAH, USA, JANUARY 21-28, 1994. JOURNAL OF CELLULAR BIOCHEMISTRY SUPPLEMENT 0 (18B). 1994. 45. ISSN: 0733-1959, XP002052633 see the whole document</p> <p>---</p> <p style="text-align: center;">-/-</p>	<p>1-3, 15-17, 22, 23, 26, 34-36</p>



Further documents are listed in the continuation of box C



Patent family members are listed in annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"8" document member of the same patent family

1

Date of the actual completion of the international search Date of mailing of the international search report

20 January 1998

03/02/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Hix, R

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/15941

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WEINREICH M D ET AL: "Evidence that the cis preference of the Tn5 transposase is caused by nonproductive multimerization." GENES & DEVELOPMENT 8 (19). 1994. 2363-2374. ISSN: 0890-9369, XP002052634 cited in the application see the whole document ---	1, 4, 5, 15, 18, 19, 22, 24, 26, 34, 37, 38
Y	DELONG, ALISON ET AL: "Trans-acting transposase mutant from Tn5" PROC. NATL. ACAD. SCI. U. S. A. (1991), 88(14), 6072-6 CODEN: PNASA6;ISSN: 0027-8424, 1991, XP002052635 see the whole document ---	1-47
Y	WIEGAND, TORSTEN W. ET AL: "Characterization of two hypertransposing Tn5 mutants" J. BACTERIOL. (1992), 174(4), 1229-39 CODEN: JOBAAY;ISSN: 0021-9193, February 1992, XP002052636 see the whole document ---	1-47
Y	WIEGAND, TORSTEN WALTER: "Transposase mutants that increase the transposition frequency of Tn5" (1993) 164 PP. AVAIL.: UNIV. MICROFILMS INT., ORDER NO. DA9315014 FROM: DISS. ABSTR. INT. B 1993, 54(6), 2886, 1993, XP002052637 see the whole document ---	1-47
Y	WEINREICH M D ET AL: "A functional analysis of the Tn5 transposase. Identification of domains required for DNA binding and multimerization." J. MOL. BIOL., vol. 241, 1993, pages 166-177, XP002052638 see the whole document ---	1-47
Y	JILK R A ET AL: "The organization of the outside end of transposon Tn5." JOURNAL OF BACTERIOLOGY, vol. 178, no. 6, March 1996, pages 1671-1679, XP002052640 see the whole document ---	1-47
	-/--	

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/15941

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	R.C. JOHNSON ET AL.: "DNA sequences at the ends of transposon Tn5 required for transposition." NATURE, vol. 304, 21 July 1983, pages 280-282, XP002052641 cited in the application see the whole document ---	1-47
P,Y	YORK, DONA ET AL: "Purification and biochemical analyses of a monomeric form of Tn5 transposase" NUCLEIC ACIDS RES. (1996), 24(19), 3790-3796 CODEN: NARHAD;ISSN: 0305-1048, 1996, XP002052642 see the whole document ---	1-47
P,X	ZHOU M ET AL: "Tn5 transposase mutants that alter DNA binding specificity." JOURNAL OF MOLECULAR BIOLOGY 271 (3). 1997. 362-373. ISSN: 0022-2836, XP002052643 see the whole document ---	1-47
P,X	YORK, DONA ET AL: "DNA binding and phasing analyses of Tn5 transposase and a monomeric variant" NUCLEIC ACIDS RES. (1997), 25(11), 2153-2160 CODEN: NARHAD;ISSN: 0305-1048, 1997, XP002052645 see the whole document -----	1-47